NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

THE NEW DATA ASSIMILATION SYSTEM AT THE ITALIAN AIR FORCE WEATHER SERVICE: DESIGN AND PRELIMINARY RESULTS

by

Massimo Bonavita

September 2002

Thesis Advisor:	Carlyle H. Wash
Second Reader:	Roger T. Williams

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** September 2002	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE The New Data Assimilation System at the Italian Air Force Weather Service: Design and Preliminary Results			**5. FUNDING NUMBERS**
6. AUTHOR(S) Massimo Bonavita			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT ***(maximum 200 words)***

A new data assimilation system has been designed and implemented at the National Center for Aeronautic Meteorology and Climatology of the Italian Air Force (CNMCA) and it is undergoing testing before eventual operational use. The new system is based on an "observation space" version of the 3D-Var method for the objective analysis component, and on the High Resolution Regional Model (H.R.M) of CNMCA for the prognostic component. New features of the system include completely rewritten correlation functions in spherical geometry, derivation of the objective analysis parameters from a statistical analysis of the innovation increments, introduction of an anisotropic component in the correlation functions, solution of analysis equations by a preconditioned conjugate gradient descent method. The analysis and forecast fields derived from the assimilation system are being subjectively and statistically evaluated through comparisons with parallel runs based on European Centre for Medium Range Weather Forecast (ECMWF): preliminary results of these studies are also presented.

14. SUBJECT TERMS Meteorology, Objective Analysis, Data Assimilation			**15. NUMBER OF PAGES** 90
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

THE NEW DATA ASSIMILATION SYSTEM AT THE ITALIAN AIR FORCE WEATHER SERVICE: DESIGN AND PRELIMINARY RESULTS

Massimo Bonavita
Captain, Italian Air Force
M.S., University of Rome "La Sapienza", 1992

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY

from the

NAVAL POSTGRADUATE SCHOOL
September 2002

Author: Massimo Bonavita

Approved by: Prof. Carlyle H. Wash
Thesis Advisor

Prof. Roger T. Williams
Second Reader/Co-Advisor

Prof. Carlyle H. Wash
Chairman, Department of Meteorology

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

A new data assimilation system has been designed and implemented at the National Center for Aeronautic Meteorology and Climatology of the Italian Air Force (CNMCA) in order to improve its numerical weather prediction capabilities and provide more accurate guidance to operational forecasters. The system, which is undergoing testing before eventual operational use, is based on an "observation space" version of the 3D-Var method for the objective analysis component, and on the High Resolution Regional Model (H.R.M) of CNMCA for the prognostic component. New features of the system include completely rewritten correlation functions in spherical geometry, derivation of the objective analysis parameters from a statistical analysis of the innovation increments, introduction of an anisotropic component in the correlation functions, solution of analysis equations by a conjugate gradient descent method. The analysis and forecast fields derived from the assimilation system are subjectively and statistically evaluated through comparisons with parallel runs based on European Centre for Medium Range Weather Forecast (ECMWF): preliminary results of these studies are also presented.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to the many people who helped him overcome the technical and bureaucratic difficulties involved in the successful completion of this project.

Deep-felt thanks go to Prof. Carlyle H. Wash, Col.(ret) Gary Roser and Ms Annamaria Munno for giving him the opportunity to attend the Naval Postgraduate School and for establishing a course of studies tailored to his specific needs.

The author is indebted to Col (ret) Maurizio Bassani, Col Massimo Capaldo and Lt. Col. Massimo Ferri for their encouragement and support of the project the present thesis is based upon; Capt. Lucio Torrisi and Ten. Antonio Vocino for their valuable help in the practical implementation of the algorithms presented in this work; Prof. Carlyle H. Wash, Prof. Roger T. Williams, Prof. Robert L. Haney and Dr. Nancy Baker for their careful review of the manuscript and many helpful suggestions.

This work is dedicated to the memory of Dr. Roger Daley, whom the author had the opportunity of meeting at the Naval Research Laboratory during a previous stay at the Naval Postgraduate School. Dr. Daley's work has been a continuous source of guidance and inspiration for the author's efforts.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Already in 1911 Bjerknes had clearly shown that the problem of weather forecasting could be thought of as an "initial condition problem" (Bjerknes, 1911). By this term we indicate the mathematical problem of predicting the future state of a physical system once the initial conditions of the system are known together with the equations governing its evolution in time.

In order to solve it successfully Bjerknes had also proposed an operative procedure based on three main components:

1. The *Observing* component;
2. The *Diagnostic* component;
3. The *Prognostic* component.

A worldwide network of in-situ and satellite-based observing systems today composes the *Observing* component. The average daily numbers of the more common observations disseminated over the Global Telecommunications System to the main weather centers are summarized in Table 1.1 (ECMWF Global Data Monitoring Report, November 2001)

This enormous quantity of data is composed of observations irregularly distributed in space and taken at different and often "asynoptic" times. The *Diagnostic* component of the forecasting system is responsible for producing an estimate of the "true" state of the atmosphere over a regular spatial grid at a given time.

Starting from this well defined initial state, the "primitive" equations describing the behavior of the atmospheric system are marched forward in time in order to produce an estimate of the state of the atmosphere at some future time (*Prognostic* component).

The main subject of the present thesis is the new *diagnostic* component of the CNMCA forecasting system: its design, implementation issues and preliminary results. The motivation for this work lies mainly in the deeply felt need to be able to take advantage of the increasing number of satellite derived observations which are nowadays

available and that cannot be easily accommodated in less recent objective analysis algorithms.

Also a brief description of the limited area model (HRM) used in the *Prognostic* component of the system will be given, together with a discussion of the main types of observations currently used and the relevant observation error statistics.

Observation type	Number of Obs.	Description
SYNOP/SHIP	53268	Surface weather observations over land or on the sea
BUOY/DRIFTER	6300	Observations from moored/ drifting buoys
TEMP	1162	Upper air radiosonde observations
TEMP/PILOT	300	Upper air wind observations
AIRCRAFT	37359	Automatic and manual observations of Temperature and wind from aircraft
SATOB	230970	Cloud motion winds from geostationary satellites imagery
NOAA 15/16 ATOVS	612488	Polar satellite derived temperature and humidity profiles

Table 1.1 Meteorological observations disseminated daily over the Global Telecommunications System (GTS)

A. ATMOSPHERIC DATA ASSIMILATION

According to a widely accepted definition (Daley, 1991) a modern data assimilation system is composed of four main components:

1. Data quality control;
2. Objective analysis;
3. Initialization of the analyzed fields;
4. Short range run of the prognostic model in order to produce an initial estimate of the atmospheric state for the successive analysis step (*First guess* or *background* fields).

A schematic representation of the 6-h intermittent data assimilation system of CNMCA is given in Fig.1.1. A brief description of components 1., 2. and 4. follows, while discussion of the proposed new realization of the objective analysis component is deferred to Chapters 2-3.

1. Data Quality Control

The data quality control step is of paramount importance in order to prevent erroneous data from being fed to the objective analysis step with deleterious results to the performance of the system. Many notable cases of recent failures of numerical weather prediction (NWP) systems to correctly forecast high impact weather conditions (1999 Christmas storm in Europe, January 2000 snow storm over the eastern coast of the US) can be attributed to the inaccuracies in the initial conditions. Often these problems can be traced back to the rejection of good observations, or the inclusion of faulty ones, by the data checking algorithms.

At CNMCA the data quality control of the observations is performed in two distinct steps. The first one, called "Observation Pre-Processing" has the purpose of assigning a degree of confidence to each reported datum. This is done through a series of checks that include:

1. Check against climatological gross limits;
2. Internal consistency checks (for example between reported and recomputed heights in TEMP messages);
3. Temporal and spatial consistency checks for observations from moving platforms (for example SHIP and DRIFTER messages).

Observations whose confidence level is below the 70% mark are discarded. More details can be found in Norris (1990).

To reduce redundancy of information and prevent possible numerical problems in the subsequent solution of the analysis equations, observations whose relative distance is less than the average grid distance of the numerical model (~55Km) are averaged together and combined in a "Super-Observation" (Lorenc, 1981).

In the second step of the data quality control, the final decision on whether to accept for ingestion the observations that survived the pre-processing step is taken. At the moment the decision is based solely on the normalized distance of the observations with respect to the background field. Although this method is statistically accurate, it can lead to rejection of good data in cases of rapidly evolving and poorly forecasted weather situations. Work is in progress on a buddy-check type of algorithm for estimating the accuracy of marginal observations.

2. Initialization

After the objective analysis step has blended in a statistically “optimum” way the information from the first guess fields and the new observations, a set of analyzed meteorological fields is produced which is suitable for a “synoptic” use (i.e. diagnosis of current state of the atmosphere for nowcasting and short range forecasting purposes), but not as initial conditions for the integration of a primitive equation model. The main reason for this lies in the fact that the imposed balance between the wind and mass observation increments are linear simplified conditions (approx. geostrophic, non-divergent), while the first guess fields implicitly satisfy the multivariate nonlinear conditions of the numerical model. As a result, the integration of non-initialized fields would cause the model to go through a geostrophic adjustment process with the excitation of inertia-gravity waves and the consequent degradation and noisiness of the forecast fields in the first 6-12 h.

To avoid these undesirable effects, the “Adiabatic Implicit Normal Mode Initialization” technique is used. A detailed explanation can be found in Temperton (1988), but the main ideas can be summarized as follows. The analyzed fields are projected over the normal modes of a linearized version of the model equations. These

normal modes can be classified (at least for the extratropics) based on their respective frequencies or propagation velocities: “fast” modes, corresponding to inertia-gravity waves, “slow” modes, corresponding to meteorological Rossby waves. The result of the projection operation are two sets of ordinary differential equations whose integration in time gives the time evolution of the amplitudes of the normal modes. The imposition of appropriate conditions (so called Machenauer conditions) on the time tendencies of the amplitudes of the normal modes leads to an effective filtering of the high-frequency modes and removal of spurious numerical noise (see pp.377-385 in Haltiner & Williams, 1980).

3. Prognostic Model

The numerical model used to produce the first guess fields in the intermittent data assimilation scheme is the High-Resolution Regional Model (HRM) of CNMCA. The HRM is a modified version of the Deutscher Wetterdienst EM/DM model (Majewski, 2001), adapted to run on Compaq Alpha servers. The main numerical features of this hydrostatic primitive equation model are summarized below:

1. Lat/Lon rotated coordinates grid, 0.5° resolution;
2. C-type Arakawa grid, 2^{nd} order centered finite difference scheme;
3. Hybrid vertical coordinate, 31 model levels;
4. Split semi-implicit time integration scheme;
5. Integration of Helmholtz equation through Fast Fourier Transform and Gauss method;
6. Davies formulation of boundary conditions;
7. 4^{th} order diffusive damping term.

As for the physics package of the model, its main characteristics are:

1. Rytter and Geleyn (1992) radiative scheme;
2. Stratiform precipitation scheme with clouds microphysics parameterization;
3. Mass flux convective scheme (Tiedtke, 1989);
4. Two-level vertical diffusion scheme (Mellor and Yamada, 1974), similarity theory at the surface (Louis, 1979);
5. Two-level soil scheme;

The operational area of model integration is shown in Fig.1.2. Embedded in this model is another version of HRM (called Med-HRM) run at double horizontal resolution (0.25° effective grid spacing) over the Mediterranean basin (Fig.1.3). Three hourly ECMWF forecast fields give the boundary conditions for the HRM model.

The interface between the objective analysis step and the prognostic model is realized through three additional software packages:

1. Insertion: spline interpolation of analyzed fields on pressure levels to hybrid coordinate model levels;
2. IFS2HRM: interpolation and adaptation of ECMWF boundary fields to HRM grid and prognostic variables;
3. Daily blending of CNMCA 12Z analysis fields with ECMWF 12Z analysis fields.

For testing purposes, the assimilation cycle is run on a Compaq DS20E server, with a 3-h data window around the analysis nominal time. The run time for an average number of independent observations (~3500) is around 45 minutes, which, considering the time necessary for the post-processing elaborations, leads to the availability of the

analyzed fields three hours after the analysis nominal time. Twice daily (at 00Z and 12 Z), an extended run (+48h) of the HRM model based on the assimilation cycle analysis is performed.

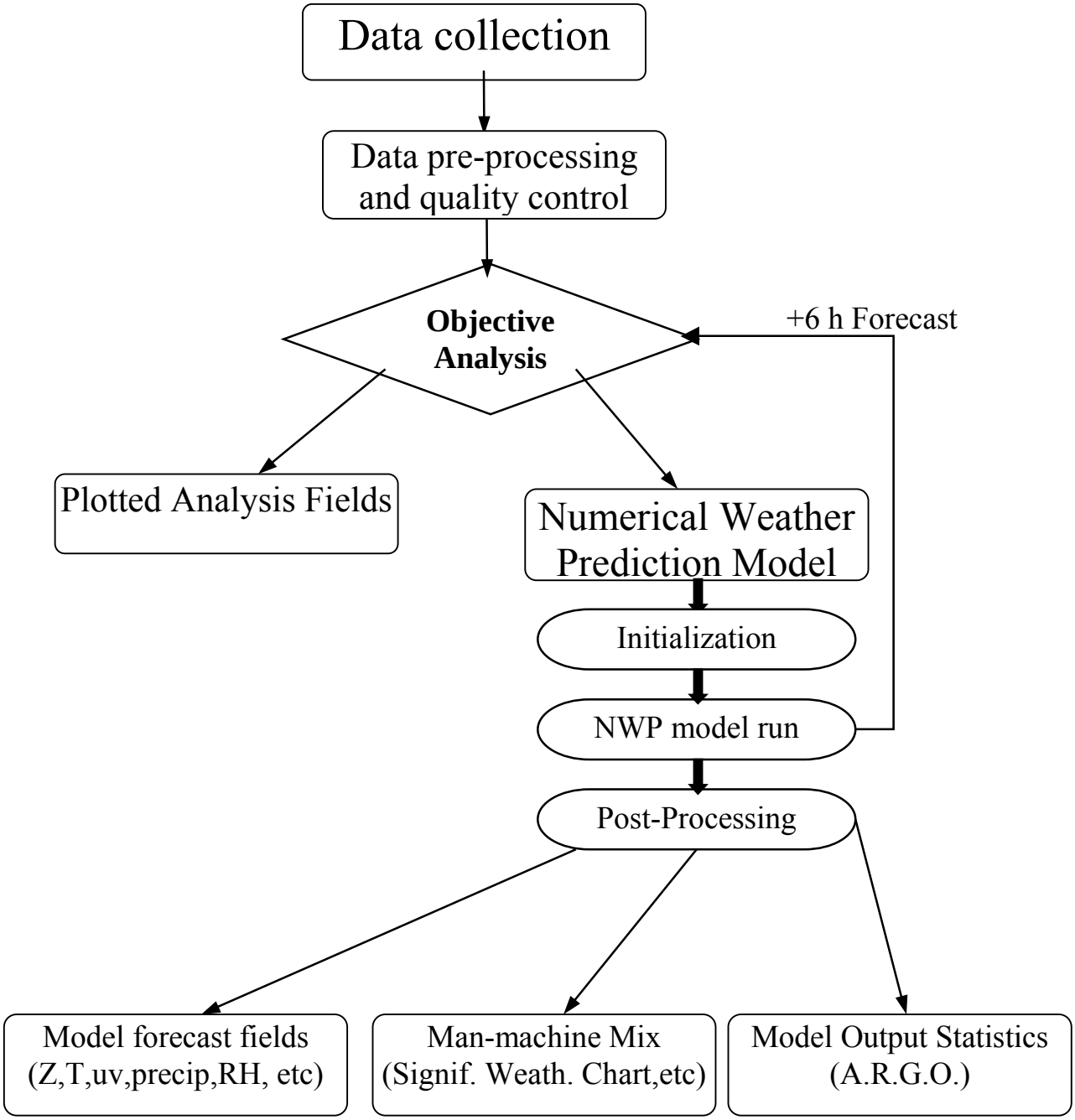

Figure 1.1 Data Assimilation Cycle at CNMCA.

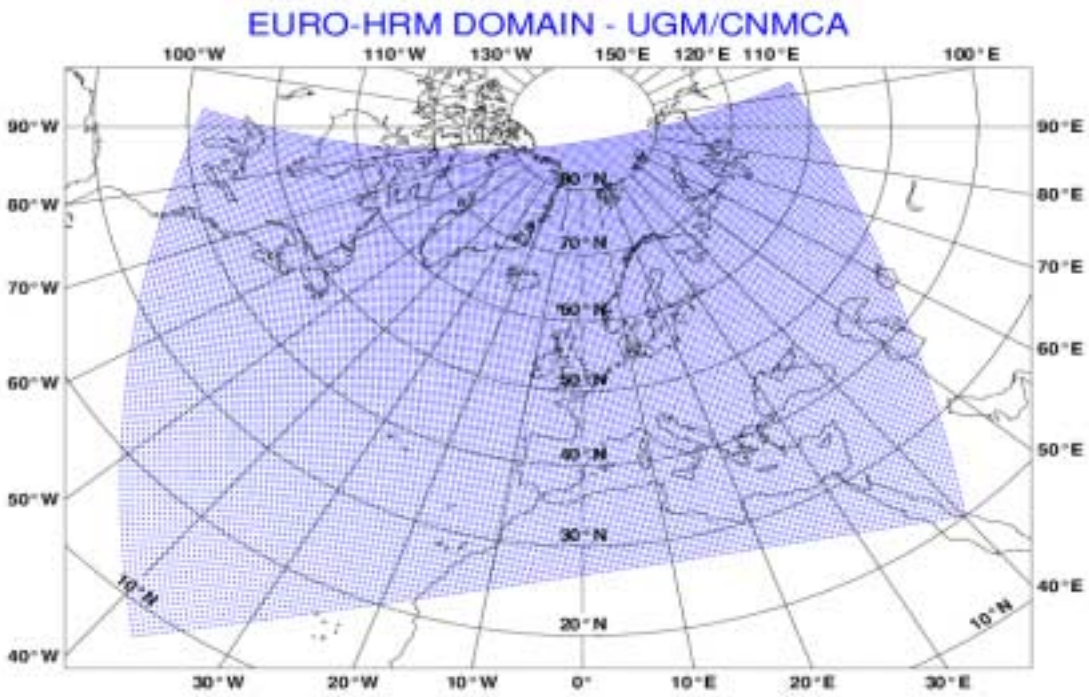

Figure 1.2 Regional model (HRM) domain of integration.

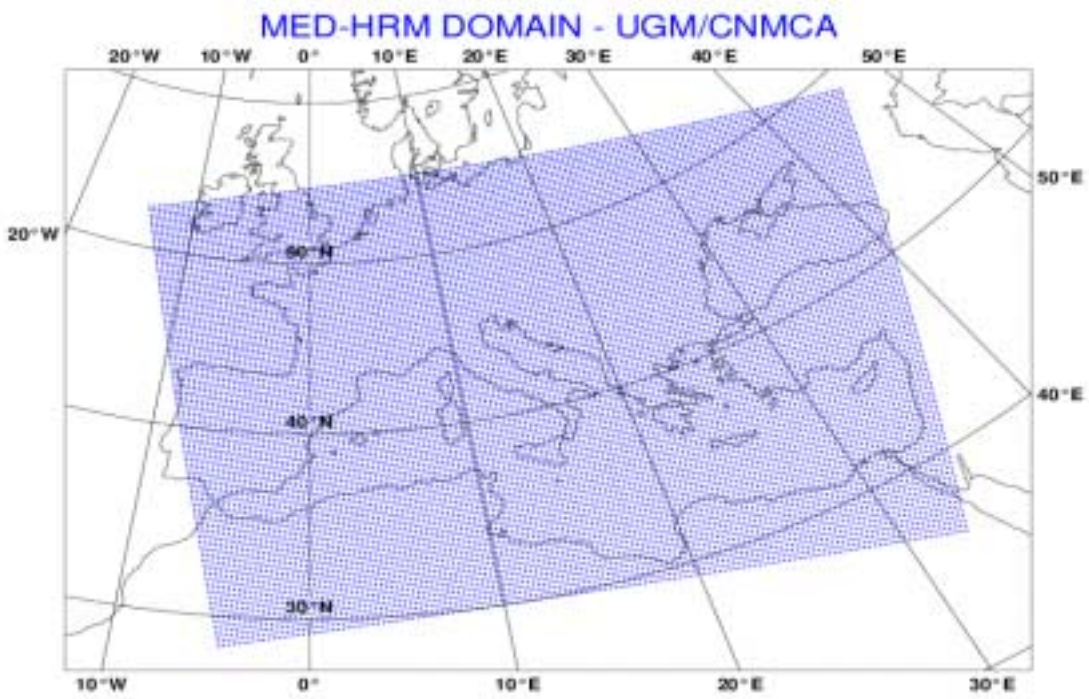

Figure 1.3 Regional model (MED-HRM) domain of integration.

THIS PAGE INTENTIONALLY LEFT BLANK

II. MULTIVARIATE VARIATIONAL DATA ASSIMILATION

Spatial objective analysis of meteorological fields can be traced back to the pioneering work of a group of research workers (J. Von Neumann, J. Charney, J. Smagorinsky and others) at Princeton Institute of Advanced Studies in the late 1940s (see Daley, 1991, for more details and a historical perspective). The early techniques were based on the concept of *function fitting:* the meteorological field to be analyzed was expanded in a finite series of known functions. The coefficients of the expansion were to be determined by a least square minimization of some function of the distance between the fitting function and the observation. The procedure leads to the solution of a linear system or, equivalently, the inversion of a square matrix (Gram matrix), which, in the case of global fitting, can become extremely large and expensive to compute. Due to this and other technical problems (possible ill-conditioning of the Gram matrix, underfitting or overfitting due to the non-stationariety of the observing system) the function fitting technique was quickly superseded in operational practice by the *method of successive corrections* (SCM) (Bergthorsson and Doos, 1955). In this technique, the function fitting is local instead of global (i.e., only the observations within a predetermined radius of the analyzed grid point influence the analysis) and the weights are specified a-priori as monotonically decreasing functions of the distance between observation and analysis grid point. Besides introducing a very efficient and robust algorithm, Bergthorsson and Doos also brought in a number of other ideas which are still in wide use in operational objective analysis schemes: the use of a forecast first guess (or background) field and the computation of the a-priori weights through a statistical analysis of the objective analysis errors. For all these attractive properties, the SCM has some important limitations, at least in its simpler formulations. It gives too much weight to the observations and it ignores the cross-correlations between observations.

Although most of the problems can be overcome with more sophisticated, iterated version of the algorithm (Bratseth, 1986), in operational practice the SCM algorithm was superseded in the 1980s by a direct method of solution of the analysis equations known as *"Statistical Interpolation"* or *"Optimal Interpolation"* (OI). The roots of the method

go back to the work of Kolmogorov (1941) and Wiener (1949) and it is interesting to note that OI, as other recent techniques, found its early applications not in meteorology but in other fields of the Earth's sciences and also in engineering. The seminal work that brought the technique to the attention of the meteorological community was that of Gandin (1963), who gave a theoretical derivation of the algorithm and discussed the main issues of its application in an operational environment.

The OI method derives its name from the fact that it tries to minimize (optimality principle) the expected variance of the analyzed fields through the a-priori knowledge of the error characteristics of both the observations and the background fields. In operational practice a number of assumptions are made which make the method sub-optimal. Apart from fundamental issues such as the imperfect knowledge of the observations and background fields error statistics, which are also often assumed stationary, homogeneous, isotropic and separable into the horizontal (or constant pressure) and the vertical components, the main approximations commonly used are:

1. Local approximation: only a limited number of observations enter into the solution of the analysis equations for any given grid point (or analysis volume, Lorenc, (1981));
2. Only observations that are linearly related to the prognostic variables of the model are commonly used, although extensions of the method to handle non-linear observations are possible (Ledvina and Pfaendtner, 1995).

In the 1990s, due to progress in computer processing speed and data storage capabilities, the variational approach to the objective analysis problem has gained popularity as an elegant and powerful mathematical formalism capable of overcoming some of the problems mentioned above (Parrish and Derber, 1992; Rabier and Courtier, 1992; Cohn et al., 1998; Daley and Barker, 2000). The methods developed by these authors are known under the general name of 3D-Var algorithms and, apart from implementation details, are basically equivalent. A scalar measure (usually called cost function) of the distance between the observations and background fields to the analysis fields is minimized with respect to the unknown analysis values, thus producing the

maximum likelihood estimate of the atmospheric state. In the case of quadratic cost functions, the minimization procedure leads to a set of linear analysis equations that are a generalization of the OI analysis equations. This is not a mere coincidence since, for stochastic variables with a Gaussian error distribution (as many weather fields can be thought of having), it can be shown (Daley, 1991) that the minimum variance estimate and the maximum likelihood estimate are coincident.

In the 3D-Var methods a basic assumption of OI is retained, namely that of the stationary character of the background error statistics. Relaxation of this hypothesis leads to various modern analysis techniques, commonly known as Four Dimensional Data Assimilation Methods (FDDA), which can be roughly divided in two main categories (Menard, 1994): time extensions of variational methods, known as 4D-Var, where the fit to the data is performed over an extended period of time (instead of intermittently) using the model as a strong constraint (i.e., strict consistency with the model, assumed perfect, is required); methods derived from estimation theory (Kalman-Bucy filter and its various modifications and extensions). In these latter methods no perfect model assumption is required; however, common to both kind of algorithms are the great computational and storage requirements that, at present, make them unfeasible options except for the largest weather centers.

In the present work we concentrate on the version of the 3D-Var algorithm which forms the basis of the new objective analysis system at CNMCA. To the author's knowledge, it was first implemented operationally at the Data Assimilation Office, NASA Goddard (Cohn et al, 1998), and more recently at Naval Research Laboratory, Monterey (Daley and Barker, 2000). The basic theory will be reviewed, then a thorough description of the background error covariance models used in the CNMCA implementation of the algorithm will be given. Finally the topic of a possible anisotropic, flow-dependent extension to the standard covariance modeling will be addressed.

A. THEORY OF 3D-VAR

The following is only a short account of the main ideas of 3D-Var which are relevant to the present work: the interested reader can find more details and a wider

theoretical background in the works of Lorenc, 1986; Daley, 1997; Cohn, 1997; Cohn et al., 1998; Daley and Barker, 2000.

Let $\boldsymbol{y}$ be the column vector containing all the p (~10^4) observations at the analysis time; $\boldsymbol{x_t}$ the column vector of the true values of the n (~10^6) state (prognostic) variables at the same time at the model grid points; $\boldsymbol{x_a}$ and $\boldsymbol{x_b}$ the analogous quantities for the analyzed and background fields; then the forecast error vector is:

$$\boldsymbol{e_b} = \boldsymbol{x_b} - \boldsymbol{x_t} \tag{2.1}$$

while the observation error vector:

$$\boldsymbol{e_o} = \boldsymbol{y} - H(\boldsymbol{x_t}) \tag{2.2}$$

where H is the observation (or forward) operator, i.e. the operator that performs the transformation from the state variables on grid points to the observed variables at the observing locations (in the case of linearly related variables it reduces to an interpolation operator). We make the usual assumptions that these error vectors, $\boldsymbol{e_b}$ and $\boldsymbol{e_o}$ have normal distribution functions with zero mean (i.e.: no bias) and are mutually uncorrelated.

From these we can now define the forecast error covariance matrix:

$$\boldsymbol{P_b} = < \boldsymbol{e_b}\, \boldsymbol{e_b}^T > \tag{2.3}$$

and the observation error covariance matrix:

$$\boldsymbol{R} = <\boldsymbol{e_o}\, \boldsymbol{e_o}^T> \tag{2.4}$$

By definition, both matrices are symmetric, positive definite, with dimensions *nxn* and *pxp* respectively.

Given the statistical assumptions made on $\boldsymbol{e_b}$ and $\boldsymbol{e_o}$, it can be shown (Daley, 1991, Chap.2) that the maximum likelihood estimate of the state of the atmospheric system is the one the minimizes the following cost function:

$$J = 0.5[\mathbf{y} - H(\mathbf{x}_a)]^T \mathbf{R}^{-1}[\mathbf{y} - H(\mathbf{x}_a)] + 0.5[\mathbf{x}_b - \mathbf{x}_a]^T \mathbf{P}_b^{-1}[\mathbf{x}_b - \mathbf{x}_a] \qquad (2.5)$$

i.e., minimize a scalar distance in L^2 of the analysis fields from both the observations and the first guess fields based on their respective perceived accuracies.

To find the minimum of the cost function (2.5), we differentiate it with respect to $\boldsymbol{x}_a$, obtaining the gradient of J:

$$\nabla J = -\mathbf{H}^T \mathbf{R}^{-1}[\mathbf{y} - H(\mathbf{x}_a)] + \mathbf{P}_b^{-1}\mathbf{x}_b \qquad (2.6)$$

where we made use of the property of bilinear symmetric forms $\partial/\partial \boldsymbol{x}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = (\mathbf{A} + \mathbf{A}^T)\mathbf{x} = 2\mathbf{A}\mathbf{x}$, and introduced the Jacobian matrix of the observation operator H:

$$\mathbf{H} \equiv \partial/\partial \boldsymbol{x}(H) \qquad (2.7)$$

A necessary condition for a minimum is obtained by setting $\nabla J = 0$:

$$-\mathbf{H}^T \mathbf{R}^{-1}[\mathbf{y} - H(\mathbf{x}_a)] + \mathbf{P}_b^{-1}\mathbf{x}_b + \mathbf{H}^T \mathbf{R}^{-1}\mathbf{H}[\mathbf{x}_a - \mathbf{x}_b] - \mathbf{H}^T \mathbf{R}^{-1}\mathbf{H}[\mathbf{x}_a - \mathbf{x}_b] = 0 \qquad (2.8)$$

where the vector $\mathbf{H}^T \mathbf{R}^{-1}\mathbf{H}[\mathbf{x}_a - \mathbf{x}_b]$ has been added and subtracted. Assuming the analysis fields to be only first order corrections to the first guess fields ("*Tangent Linear*" approximation) we have:

$$H(\mathbf{x}_a) = H(\mathbf{x}_b - (\mathbf{x}_a - \mathbf{x}_b)) \cong H(\mathbf{x}_b) - \mathbf{H}(\mathbf{x}_a - \mathbf{x}_b) \qquad (2.9)$$

which, when substituted into (2.8), yields:

$$- \boldsymbol{H}^T\boldsymbol{R}^{-1}[\boldsymbol{y} - H(\boldsymbol{x}_b)] + (\boldsymbol{P}_b^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})\,(\boldsymbol{x}_a - \boldsymbol{x}_b) = 0 \qquad (2.10)$$

or,

$$\boldsymbol{x}_a - \boldsymbol{x}_b = (\boldsymbol{P}_b^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\,\boldsymbol{H}^T\boldsymbol{R}^{-1}[\boldsymbol{y} - H(\boldsymbol{x}_b)] \qquad (2.11)$$

This form of the analysis equations involves covariance matrices computed in the grid space of the model (or, equivalently, in its spectral space), which, in the present case, would involve matrices of order $\sim 10^6 \text{x} 10^6$ (although, in spectral space and under the assumptions of homogeneity and horizontal isotropy for correlations, they can be shown to be diagonal or block diagonal: Courtier, 1997; Courtier at al, 1998).

A less expensive but equivalent form of the analysis equations can be derived making use of the Sherman-Morrison-Woodbury formula (for the details of the derivation see Courtier, 1997), or, more simply by noting that:

$$\boldsymbol{H}^T\boldsymbol{R}^{-1}\,(\boldsymbol{P}_b^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1} = (\boldsymbol{H}\,\boldsymbol{P}_b\,\boldsymbol{H}^T + \boldsymbol{R})\,\boldsymbol{P}_b\,\boldsymbol{H}^T \qquad (2.12)$$

and that all the matrices considered and their inverse are positive definite. Inserting 2.12 into 2.11 we find:

$$\boldsymbol{x}_a - \boldsymbol{x}_b = \boldsymbol{P}_b\,\boldsymbol{H}^T\,(\boldsymbol{H}\,\boldsymbol{P}_b\,\boldsymbol{H}^T + \boldsymbol{R})^{-1}[\boldsymbol{y} - H(\boldsymbol{x}_b)] \qquad (2.13)$$

This formulation is the one used in the present work. To the author's knowledge, it was first implemented in the NASA/Goddard Data Assimilation Office "Physical Space Statistical Analysis System" (Cohn et al., 1998) and, more recently, in the NAVDAS assimilation system of NRL Monterey (Daley and Barker, 2000).

From inspection it is clear that we are dealing with an observation space algorithm: $\boldsymbol{R}$ is defined in observation space while $\boldsymbol{P}_b$ is projected into it by means of the observation operator $\boldsymbol{H}$. Taking into account the average number of observations currently entering into the CNMCA assimilation system ($\sim 10^4$), we can see that this

approach is able to considerably reduce the computational and storage load on the computing facilities without placing any constraint on the covariance function models.

The main differences with respect to standard OI algorithms are twofold: first there is the explicit use of the observation operator *H*, which will provide a natural way to extend the objective analysis to include observations related in a nonlinear way to the prognostic variables of the model (see below); secondly, as it will be shown below, the solution to (2.12) is computed globally (i.e.: making use of all available observations for every grid point), without the need for data selection procedures.

The extension of (2.12) to non-linear observations (such as radiances, wind speeds, total column water content, etc., where $\boldsymbol{H}$ depends on the model state $\boldsymbol{x}$) is conceptually simple: we define a series of system states:

$$\boldsymbol{x_0}=\boldsymbol{x_b}, \boldsymbol{x_1}, \boldsymbol{x_2},\ldots,\boldsymbol{x_{i-1}},\boldsymbol{x_i},\boldsymbol{x_{i+1}},\ldots \tag{2.14}$$

where the starting state coincides with the first guess fields.

Then, defining the Jacobian of the observation operator around the *i* state of the system

$$\boldsymbol{H_i} \equiv \partial/\partial\boldsymbol{x}(H)_{x=x_i} \tag{2.15}$$

the nonlinear version of the analysis equations is found by looking for a Newtonian iterative solution to the problem of finding the root of the Jacobian of the cost function ($\nabla J=0$; eq. 2.8) : starting from a first guess state $\boldsymbol{x_0}=\boldsymbol{x_b}$, the $i+1$ state is given by:

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \nabla^2 J(\boldsymbol{x}_i)^{-1} \cdot \nabla J(\boldsymbol{x}_i) \tag{2.16}$$

where $\nabla^2 J(\boldsymbol{x}_i)^{-1} = (\boldsymbol{P}_b^{-1} + \boldsymbol{H}^T_i \boldsymbol{R}^{-1} \boldsymbol{H}_i)$ is the Hessian matrix of the cost function *J* and it can be shown to be the inverse of the analysis error covariance matrix.

After some manipulation, 2.16 can be recast in the computationally more efficient form:

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_0 + \boldsymbol{P}_b \boldsymbol{H}^T{}_i \, (\boldsymbol{H}_i \boldsymbol{P}_b \boldsymbol{H}^T{}_i + \boldsymbol{R})^{-1} [\boldsymbol{y} - H(\boldsymbol{x}_i) + \boldsymbol{H}_i(\boldsymbol{x}_i - \boldsymbol{x}_0)] \tag{2.17}$$

The iterative procedure stops when the difference between $\boldsymbol{x}_{i+1}$ and $\boldsymbol{x}_i$ becomes smaller than a predefined value. Unlike the linearized version of the analysis equations, which result from a quadratic cost function, the minimization procedure (2.15) cannot be guaranteed to converge and it can have more the one minimum. Up to now only observations linearly related to the prognostic variables of the model have been used in the experimental phase.

B. IMPLEMENTATION ISSUES

The practical implementation of objective analysis equations (2.13) has been carried out as follows. The set of equations (2.13) can be solved in two steps. First solution of the linear *pxp* system:

$$(\boldsymbol{H} \boldsymbol{P}_b \boldsymbol{H}^T + \boldsymbol{R}) \, \boldsymbol{z} = \boldsymbol{y} - H(\boldsymbol{x}_b) \tag{2.18}$$

in the unknown vector $\boldsymbol{z}$. Secondly, projection of the solution on grid space via:

$$\boldsymbol{x}_a - \boldsymbol{x}_b = \boldsymbol{P}_b \boldsymbol{H}^T \boldsymbol{z} \tag{2.19}$$

The second step amounts to perform a matrix-vector product between the *pxn* matrix $\boldsymbol{P}_b \boldsymbol{H}^T$ and the *p* vector $\boldsymbol{z}$, which can be computed efficiently in Fortran90 code.

The first step involves the solution of a large, sparse, symmetric and definite positive linear system. It can be shown (Golub and van Loan, 1996) that this step is mathematically equivalent to finding the minimum of the following cost function:

$$F(\boldsymbol{z}) = 1/2 \boldsymbol{z}^T (\boldsymbol{H} \boldsymbol{P}_b \boldsymbol{H}^T + \boldsymbol{R}) \boldsymbol{z} - \boldsymbol{z}^T (\boldsymbol{y} - H(\boldsymbol{x}_b)) \tag{2.20}$$

which can be done using a standard Conjugate Gradient (CG) Algorithm (Subroutine F11GBZ, NAG Library, Mark 17,1995, has been used).

In numerical experiments, it has been found convenient to implement a scaled version of 2.13. This is due to the reason that, in this way, the condition number of the matrix $(\boldsymbol{H}\,\boldsymbol{P}_b\,\boldsymbol{H}^T + \boldsymbol{R})$ (defined as the absolute value of the ratio of the maximum to the minimum eigenvalue) shows a considerable decrease, thus speeding up the convergence of the descent algorithm.

The scaled version of (2.13) is found following Daley and Barker (2000). If we redefine the observation operator $\boldsymbol{H}$ as the product of a spatial interpolation operator $\boldsymbol{H}_*$ and a real observation operator $\boldsymbol{H}$ (i.e. $\boldsymbol{H} \rightarrow \boldsymbol{H}_*\,\boldsymbol{H}$) and set:

$$\boldsymbol{P}_b\,\boldsymbol{H}_*^{\boldsymbol{T}} = \boldsymbol{P}_b^{gr/ob} \quad \text{and,} \quad \boldsymbol{H}_*\,\boldsymbol{P}_b\,\boldsymbol{H}_*^{\boldsymbol{T}} = \boldsymbol{P}_b^{ob/ob} \tag{2.21}$$

$$\boldsymbol{P}_b^{gr/ob}\boldsymbol{H}^{\boldsymbol{T}} = \boldsymbol{S}_{\boldsymbol{b}}^{1/2}\,\boldsymbol{C}_b^{gr/ob}\,[\boldsymbol{S}_{\boldsymbol{b}}^{\boldsymbol{ob}}]^{1/2}\boldsymbol{H}^{\boldsymbol{T}} \tag{2.22}$$

$$\boldsymbol{H}\boldsymbol{P}_b^{ob/ob}\boldsymbol{H}^{\boldsymbol{T}} = \boldsymbol{H}[\boldsymbol{S}_{\boldsymbol{b}}^{\boldsymbol{ob}}]^{1/2}\,\boldsymbol{C}_b^{ob/ob}\,[\boldsymbol{S}_{\boldsymbol{b}}^{\boldsymbol{ob}}]^{1/2}\boldsymbol{H}^{\boldsymbol{T}} \tag{2.23}$$

$$\boldsymbol{S}_{\boldsymbol{h}} = diag(\boldsymbol{H}_*\,\boldsymbol{P}_b^{ob/ob}\,\boldsymbol{H}_*^{\boldsymbol{T}}) \tag{2.24}$$

Where $\boldsymbol{S}_{\boldsymbol{b}} = diag(\boldsymbol{P}_b)$, (i.e. the background variances of the analysis variables), $\boldsymbol{S}_{\boldsymbol{b}}^{\boldsymbol{ob}} = diag(\boldsymbol{P}_b^{ob})$, (i.e. the observation variances of the analysis variables), and $\boldsymbol{C}_b^{gr/ob}$, $\boldsymbol{C}_b^{ob/ob}$ are the corresponding correlation matrices., then the scaled form of 2.13 is:

$$\boldsymbol{x}_a - \boldsymbol{x}_b = \boldsymbol{S}_{\boldsymbol{b}}^{1/2}\,\boldsymbol{C}_b^{gr/ob}\,[\boldsymbol{S}_{\boldsymbol{b}}^{\boldsymbol{ob}}]^{1/2}\boldsymbol{H}^{\boldsymbol{T}}\,\boldsymbol{S}_{\boldsymbol{h}}^{-1/2}[\boldsymbol{C}_h^{ob/ob} + \boldsymbol{S}_{\boldsymbol{h}}^{-1/2}\boldsymbol{R}\,\boldsymbol{S}_{\boldsymbol{h}}^{-1/2}]\,\boldsymbol{S}_{\boldsymbol{h}}^{-1/2}[\boldsymbol{y} - H(\boldsymbol{x}_b)] \tag{2.25}$$

and $\boldsymbol{C}_h^{ob/ob} = \boldsymbol{S}_{\boldsymbol{h}}^{-1/2}\boldsymbol{H}[\boldsymbol{S}_{\boldsymbol{b}}^{\boldsymbol{ob}}]^{1/2}\,\boldsymbol{C}_b^{ob/ob}[\boldsymbol{S}_{\boldsymbol{b}}^{\boldsymbol{ob}}]^{1/2}\boldsymbol{H}^{\boldsymbol{T}}\,\boldsymbol{S}_{\boldsymbol{h}}^{-1/2}$.

The main point here is the rescaling of the $(\boldsymbol{H}\,\boldsymbol{P}_b\,\boldsymbol{H}^{\boldsymbol{T}} + \boldsymbol{R})$ covariance matrix in the non-dimensional form $[\boldsymbol{C}_h^{ob/ob} + \boldsymbol{S}_{\boldsymbol{h}}^{-1/2}\boldsymbol{R}\,\boldsymbol{S}_{\boldsymbol{h}}^{-1/2}]$.

In experimental runs, without preconditioning, for an average number of observations (p~3500), the cost function (2.20) converges to within machine precision in around 150 steps (average time ~90 seconds on Compaq DS20E server). Since the computational cost of the algorithm scales as number of CG iterations multiplied by p^2, it is important, in view of ingesting asynoptic type of observations, to implement an

effective pre-conditioner for the $(\boldsymbol{H}\,\boldsymbol{P}_b\,\boldsymbol{H}^T + \boldsymbol{R})$ matrix. Work in this direction is still in progress.

C. BACKGROUND ERROR COVARIANCES

"*The most important element in the statistical interpolation algorithm is the background error covariance matrix. To a large extent, the form of this matrix governs the resulting objective analysis…"* (Daley, 1991). Much effort has so far gone into the redefinition of the background error covariance model. The derivation follows the main ideas of Bergman (1979), Thiebaux et al. (1986), Thiebaux at al. (1990), Tillmann (1999). An original contribution has been the explicit formulation of the temperature-wind cross correlations in spherical coordinates, using the thermal wind relationship as a constraint. This, to the author's knowledge, is not to be found in the relevant literature.

The need to express the auto and cross-correlations in spherical coordinates instead of using a local plane projection arises from the fact that, for the correlation model used, one has non negligible correlation values at distances comparable to the Earth's radius.

Taking into account the narrower structure shown by the background vertical correlations of temperature with respect to those of geopotential and the weaker vertical correlation of radiosondes' temperature measurements with respect to geopotential observations, the choice was made to adopt the following analysis variables: temperature (T), zonal (u) and meridional (v) components of wind, surface pressure (SP) and relative humidity (RH).

1. Upper Air Analysis Covariance Model

The upper air analysis is multivariate in (T, u, v). The covariance model has been derived as follows.

Starting from the geostrophic constraint:

$$\boldsymbol{u} = -(k/\sin\varphi)\, \partial\boldsymbol{\Phi}/\partial\varphi \tag{2.26}$$

$$\boldsymbol{v} = k/(\sin\varphi\cos\varphi)\, \partial\boldsymbol{\Phi}/\partial\lambda \tag{2.27}$$

where (φ, λ) are the latitude-longitude coordinates, $\boldsymbol{\Phi}$ is the geopotential, and $k \equiv \mu/2\Omega$ is a constant that takes into account the geostrophic coupling parameter μ and the orbital angular velocity of the Earth $\boldsymbol{\Omega}$. Making use of the equation of state for dry air ($\boldsymbol{p}=\boldsymbol{\rho} R_d \boldsymbol{T}$) and the hydrostatic equation ($\partial\boldsymbol{p}/\partial\boldsymbol{\Phi}=-\boldsymbol{\rho}$), (2.25-26) can be recast as a form of the thermal wind constraint:

$$\partial\boldsymbol{u}/\partial\boldsymbol{p} = kR_d/(\boldsymbol{p}\sin\varphi)\, \partial\boldsymbol{T}/\partial\varphi \tag{2.28}$$

$$\partial\boldsymbol{v}/\partial\boldsymbol{p} = -k/(\boldsymbol{p}\sin\varphi\cos\varphi)\, \partial\boldsymbol{T}/\partial\lambda \tag{2.29}$$

Under the usual hypothesis of homogeneity, isotropy and separability for the temperature autocorrelation function, we assume the following functional form for the temperature covariance:

$$Cov(\boldsymbol{T}_i, \boldsymbol{T}_j) = \sigma_{\boldsymbol{T}}^2\, R(\tau)\, \chi^{\boldsymbol{TT}}(\boldsymbol{p}_i, \boldsymbol{p}_j) \tag{2.30}$$

where $\sigma_{\boldsymbol{T}}^2$ is the background error temperature variance (which can vary both in latitude and in pressure level), $R(\tau)$ is the quasi-horizontal (isobaric) component of the correlation model (function only of the Great Circle distance $\tau = \cos^{-1}(\sin\varphi_i \sin\varphi_i + \cos\varphi_i \cos\varphi_j \cos|\Delta\lambda|)$ of points i,j) and $\chi^{\boldsymbol{TT}}$ is the vertical part of the correlation function. Following Thiebaux et al. (1986) and Daley and Barker (2000), the functional representation for $R(\tau)$ has been chosen as a Second Order Autoregressive (SOAR) Function of the form:

$$R(\tau) = (1+c\,\tau)\exp(-c\,\tau) \tag{2.31}$$

Where the length scale c^{-1} will be specified through a statististical analysis of the observed minus forecast increments, as will be shown in Chapter 3.

The derivation that follows is, however, independent of the specific form of either $R(\tau)$ or χ^{TT}. Making use of (2.27-28) is easily seen that:

$$\partial/\partial \ln \mathbf{p_i}\ Cov(\mathbf{u}_i, \mathbf{T}_j) = \sigma_T^2\ kR_d/\sin\varphi_i\ dR(\tau)/d\tau\ \partial\tau/\partial\varphi_i\ \chi^{TT}(\mathbf{p}_i,\mathbf{p}_j) \quad (2.32)$$

$$\partial/\partial \ln \mathbf{p_j}\ Cov(\mathbf{T}_i, \mathbf{u}_j) = \sigma_T^2\ kR_d/\sin\varphi_j\ dR(\tau)/d\tau\ \partial\tau/\partial\varphi_j\ \chi^{TT}(\mathbf{p}_i,\mathbf{p}_j) \quad (2.33)$$

$$\partial/\partial \ln \mathbf{p_i}\ Cov(\mathbf{v}_i, \mathbf{T}_j) = -\ \sigma_T^2\ kR_d/(\sin\varphi_i\cos\varphi_i)\ dR(\tau)/d\tau\ \partial\tau/\partial\lambda_i\ \chi^{TT}(\mathbf{p}_i,\mathbf{p}_j) \quad (2.34)$$

$$\partial/\partial \ln \mathbf{p_j}\ Cov(\mathbf{T}_i, \mathbf{v}_j) = -\ \sigma_T^2\ kR_d/(\sin\varphi_j\cos\varphi_j)\ dR(\tau)/d\tau\ \partial\tau/\partial\lambda_j\ \chi^{TT}(\mathbf{p}_i,\mathbf{p}_j) \quad (2.35)$$

$$\partial^2/(\partial \ln \mathbf{p_i}\partial \ln \mathbf{p_j})\ Cov(\mathbf{u}_i, \mathbf{u}_j) = (\sigma_T\ kR_d)^2/(\sin\varphi_i\sin\varphi_i)\ (d^2R(\tau)/d\tau^2\ \partial\tau/\partial\varphi_i\partial\tau/\partial\varphi_j + dR(\tau)/d\tau\ \partial^2\tau/\partial\varphi_i\partial\varphi_j)\ \chi^{TT}(\mathbf{p}_i,\mathbf{p}_j) \quad (2.36)$$

$$\partial^2/(\partial \ln \mathbf{p_i}\partial \ln \mathbf{p_j})\ Cov(\mathbf{v}_i, \mathbf{v}_j) = (\sigma_T\ kR_d)^2/(\sin\varphi_i\cos\varphi_i\sin\varphi_i\cos\varphi_j)\ (d^2R(\tau)/d\tau^2\ \partial\tau/\partial\lambda_i\partial\tau/\partial\lambda_j + dR(\tau)/d\tau\ \partial^2\tau/\partial\lambda_i\partial\lambda_j)\ \chi^{TT}(\mathbf{p}_i,\mathbf{p}_j) \quad (2.37)$$

$$\partial^2/(\partial \ln \mathbf{p_i}\partial \ln \mathbf{p_j})\ Cov(\mathbf{u}_i, \mathbf{v}_j) = -\ (\sigma_T\ kR_d)^2/(\sin\varphi_i\sin\varphi_i\cos\varphi_j)\ (d^2R(\tau)/d\tau^2\ \partial\tau/\partial\varphi_i\partial\tau/\partial\lambda_j + dR(\tau)/d\tau\ \partial^2\tau/\partial\varphi_i\partial\lambda_j)\ \chi^{TT}(\mathbf{p}_i,\mathbf{p}_j) \quad (2.38)$$

$$\partial^2/(\partial \ln \mathbf{p_i}\partial \ln \mathbf{p_j})\ Cov(\mathbf{v}_i, \mathbf{u}_j) = -\ (\sigma_T\ kR_d)^2/(\sin\varphi_i\sin\varphi_i\cos\varphi_j)\ (d^2R(\tau)/d\tau^2\ \partial\tau/\partial\varphi_j\partial\tau/\partial\lambda_i + dR(\tau)/d\tau\ \partial^2\tau/\partial\varphi_j\partial\lambda_i)\ \chi^{TT}(\mathbf{p}_i,\mathbf{p}_j) \quad (2.39)$$

Similarly to (2.29) the other covariances can be written as:

$$Cov(\mathbf{u}_i, \mathbf{T}_j) = \sigma_u\ \sigma_T\ R^{uT}(\mathbf{r}_i,\mathbf{r}_j)\ \chi^{uT}(\mathbf{p}_i,\mathbf{p}_j) \quad (2.40)$$

$$Cov(\mathbf{v}_i, \mathbf{T}_j) = \sigma_v\ \sigma_T\ R^{vT}(\mathbf{r}_i,\mathbf{r}_j)\ \chi^{vT}(\mathbf{p}_i,\mathbf{p}_j) \quad (2.41)$$

$$Cov(\mathbf{u}_i, \mathbf{u}_j) = \sigma_u^2\ R^{uu}(\mathbf{r}_i,\mathbf{r}_j)\ \chi^{uu}(\mathbf{p}_i,\mathbf{p}_j) \quad (2.42)$$

$$Cov(\mathbf{v}_i, \mathbf{v}_j) = \sigma_v^2\ R^{vv}(\mathbf{r}_i,\mathbf{r}_j)\ \chi^{vv}(\mathbf{p}_i,\mathbf{p}_j) \quad (2.43)$$

$$Cov(\mathbf{u}_i, \mathbf{v}_j) = \sigma_u\ \sigma_v\ R^{uv}(\mathbf{r}_i,\mathbf{r}_j)\ \chi^{uv}(\mathbf{p}_i,\mathbf{p}_j) \quad (2.44)$$

and substituted into (2.25-32), thus obtaining for the vertical components:

$$\chi^{uu}(\mathbf{p}_i,\mathbf{p}_j) = \chi^{vv}(\mathbf{p}_i,\mathbf{p}_j) = \chi^{uv}(\mathbf{p}_i,\mathbf{p}_j) = \chi^{vu}(\mathbf{p}_i,\mathbf{p}_j) \quad (2.45)$$

$$\chi^{uT}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \chi^{vT}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \partial/\partial \ln \boldsymbol{p}_j \chi^{uu}(\boldsymbol{p}_i,\boldsymbol{p}_j) \tag{2.46}$$

$$\chi^{Tu}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \chi^{Tv}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \partial/\partial \ln \boldsymbol{p}_i \chi^{uu}(\boldsymbol{p}_i,\boldsymbol{p}_j) \tag{2.47}$$

$$\chi^{TT}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \partial^2/(\partial \ln \boldsymbol{p}_i \partial \ln \boldsymbol{p}_j)\, \chi^{uu}(\boldsymbol{p}_i,\boldsymbol{p}_j) \tag{2.48}$$

while for the isobaric components we find:

$$\sigma_u R^{uT}(\boldsymbol{r}_i,\boldsymbol{r}_j) = \sigma_T kR_d/\sin\varphi_i \; dR(\tau)/d\tau \; \partial\tau/\partial\varphi_i \tag{2.49}$$

$$\sigma_v R^{vT}(\boldsymbol{r}_i,\boldsymbol{r}_j) = -\sigma_T kR_d/(\sin\varphi_i \cos\varphi_i)\; dR(\tau)/d\tau \; \partial\tau/\partial\lambda_i \tag{2.50}$$

$$\sigma_u^2 R^{uu}(\boldsymbol{r}_i,\boldsymbol{r}_j) = (\sigma_T kR_d)^2/(\sin\varphi_i \sin\varphi_i)\; (d^2R(\tau)/d\tau^2\, \partial\tau/\partial\varphi_i \partial\tau/\partial\varphi_j + dR(\tau)/d\tau \partial^2\tau/\partial\varphi_i\partial\varphi_j) \tag{2.51}$$

$$\sigma_u \sigma_v R^{uv}(\boldsymbol{r}_i,\boldsymbol{r}_j) = -(\sigma_T kR_d)^2/(\sin\varphi_i \sin\varphi_i \cos\varphi_j)\; (d^2R(\tau)/d\tau^2\, \partial\tau/\partial\varphi_i \partial\tau/\partial\lambda_j + dR(\tau)/d\tau \partial^2\tau/\partial\varphi_i\partial\lambda_j) \tag{2.52}$$

$$\sigma_v^2 R^{vv}(\boldsymbol{r}_i,\boldsymbol{r}_j) = (\sigma_T kR_d)^2/(\sin\varphi_i\cos\varphi_i\sin\varphi_i \cos\varphi_j)\; (d^2R(\tau)/d\tau^2\, \partial\tau/\partial\lambda_i \partial\tau/\partial\lambda_j + dR(\tau)/d\tau \partial^2\tau/\partial\lambda_i\partial\lambda_j) \tag{2.53}$$

Setting

$$Lim_{(\tau \to 0)}\, \tau^{-1}\, dR(\tau)/d\tau \equiv L \tag{2.54}$$

we obtain for the geostrophically constrained wind variances:

$$Lim_{(\tau \to 0)}\, Cov(\boldsymbol{u}_i,\boldsymbol{u}_j) = \sigma_u^2 = Lim_{(\tau \to 0)}\, Cov(\boldsymbol{v}_i,\boldsymbol{v}_j) = \sigma_v^2 = -(\sigma_T kR_d/\sin\varphi_i)^2 L \tag{2.55}$$

Substituting (2.54) into (2.48-52) we completely determine the isobaric correlations:

$$R^{uT}(\boldsymbol{r}_i,\boldsymbol{r}_j) = dR(\tau)/d\tau \; \partial\tau/\partial\varphi_i \; (-L)^{-1/2} \tag{2.56}$$

$$R^{vT}(\boldsymbol{r}_i,\boldsymbol{r}_j) = -(\cos\varphi_i)^{-1}\, dR(\tau)/d\tau \; \partial\tau/\partial\lambda_i \; (-L)^{-1/2} \tag{2.57}$$

$$R^{uu}(\boldsymbol{r}_i,\boldsymbol{r}_j) = (d^2R(\tau)/d\tau^2\, \partial\tau/\partial\varphi_i \partial\tau/\partial\varphi_j + dR(\tau)/d\tau \partial^2\tau/\partial\varphi_i\partial\varphi_j)\; (-L)^{-1} \tag{2.58}$$

$$R^{uv}(\mathbf{r}_i,\mathbf{r}_j) = -\,(d^2R(\tau)/d\tau^2\,\partial\tau/\partial\varphi_i\partial\tau/\partial\lambda_j + dR(\tau)/d\tau\partial^2\tau/\partial\varphi_i\partial\lambda_j)\,(-L\cos\varphi_j)^{-1} \quad (2.59)$$

$$R^{vv}(\mathbf{r}_i,\mathbf{r}_j) = (d^2R(\tau)/d\tau^2\,\partial\tau/\partial\lambda_i\partial\tau/\partial\lambda_j + dR(\tau)/d\tau\partial^2\tau/\partial\lambda_i\partial\lambda_j)\,(-L\cos\varphi_i\cos\varphi_j)^{-1} \quad (2.60)$$

The above formulas are independent of the chosen correlation model. In the present case, assuming (2.29) as the functional representation for the isobaric temperature autocorrelation model, we obtain:

$$R^{uT}(\mathbf{r}_i,\mathbf{r}_j) = c(\tau/\sin\tau)\exp(-c\tau)(\cos\varphi_i\sin\varphi_j - \cos\varphi_j\sin\varphi_i\cos|\Delta\lambda|) \quad (2.61)$$

$$R^{Tu}(\mathbf{r}_i,\mathbf{r}_j) = -R^{uT}(\mathbf{r}_i,\mathbf{r}_j) \quad (2.62)$$

$$R^{vT}(\mathbf{r}_i,\mathbf{r}_j) = c(\tau/\sin\tau)\exp(-c\tau)(\cos\varphi_j\sin(\lambda_i-\lambda_j)) \quad (2.63)$$

$$R^{Tv}(\mathbf{r}_i,\mathbf{r}_j) = -R^{vT}(\mathbf{r}_i,\mathbf{r}_j) \quad (2.64)$$

$$R^{uu}(\mathbf{r}_i,\mathbf{r}_j) = -((\sin\tau(1-c\tau)-\tau\cos\tau)/(\sin^3\tau)\,(\cos\varphi_i\sin\varphi_j - \cos\varphi_j\sin\varphi_i\cos|\Delta\lambda|)\,(\cos\varphi_j\sin\varphi_i - \cos\varphi_i\sin\varphi_j\cos|\Delta\lambda|) - (\tau/\sin\tau)\,(\cos\varphi_i\sin\varphi_j - \cos\varphi_j\sin\varphi_i\cos|\Delta\lambda|))\exp(-c\tau) \quad (2.65)$$

$$R^{vv}(\mathbf{r}_i,\mathbf{r}_j) = ((\sin\tau(1-c\tau)-\tau\cos\tau)/(\sin^3\tau)\,(\cos\varphi_i\cos\varphi_j\sin^2|\Delta\lambda|) + (\tau/\sin\tau)\cos|\Delta\lambda|))\exp(-c\tau) \quad (2.66)$$

$$R^{uv}(\mathbf{r}_i,\mathbf{r}_j) = ((\sin\tau(1-c\tau)-\tau\cos\tau)/(\sin^3\tau)\,(\cos\varphi_i\sin\varphi_j - \cos\varphi_j\sin\varphi_i\cos|\Delta\lambda|)\,(\cos\varphi_i\sin(\lambda_i-\lambda_j) + (\tau/\sin\tau)\,(\sin\varphi_i\sin(\lambda_i-\lambda_j))\exp(-c\tau) \quad (2.67)$$

The above isobaric correlations are shown in Fig.2.1 through 2.6.

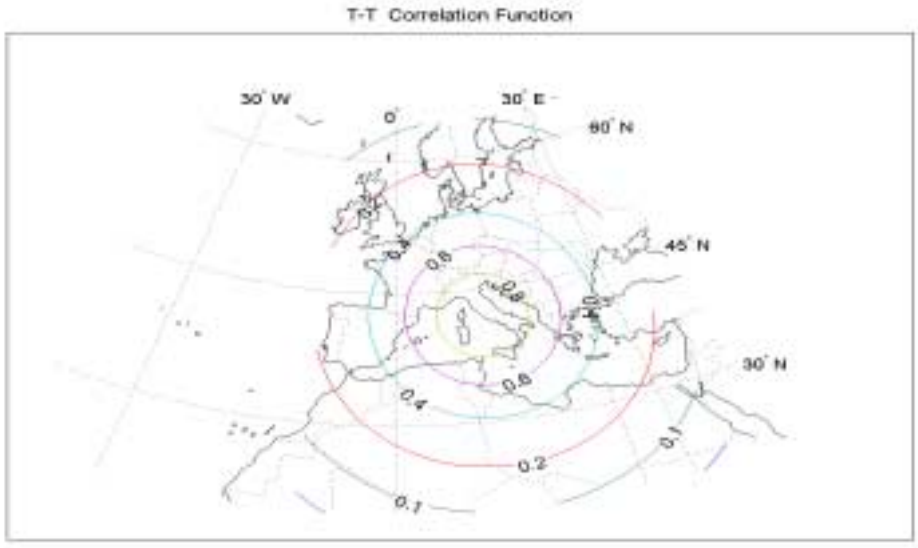

Figure 2.1 T-T correlation function (SOAR Model, *c = 0.1 rad*).

.

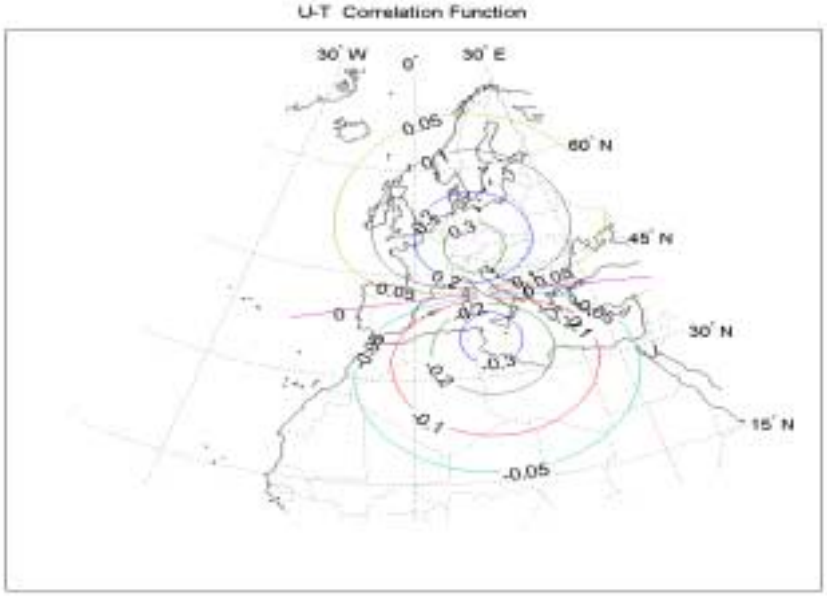

Figure 2.2 U-T correlation function (SOAR Model, *c = 0.1 rad*).

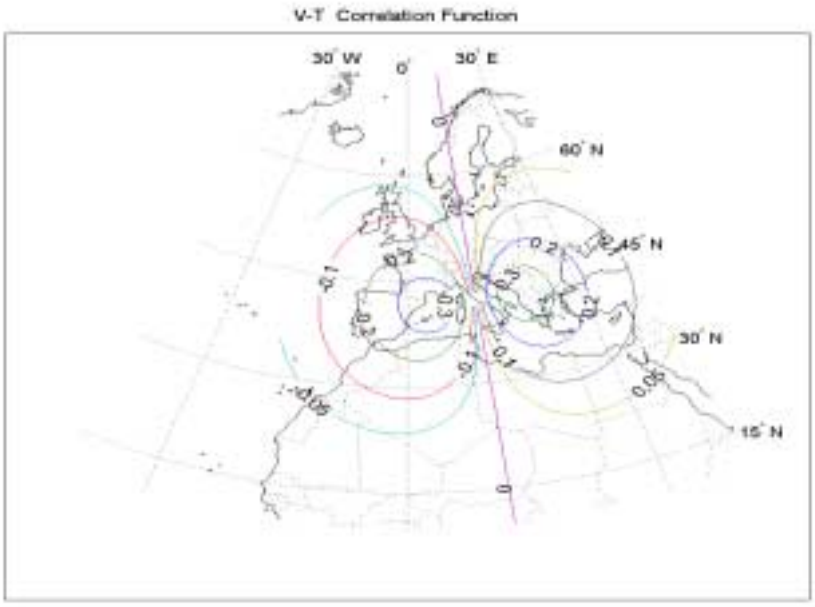

Figure 2.3 V-T correlation function (SOAR Model, *c = 0.1 rad*).

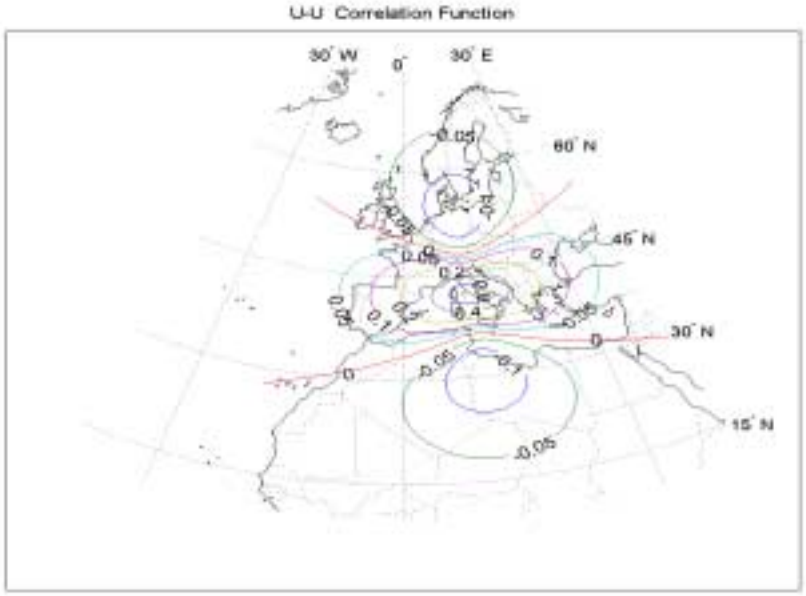

Figure 2.4 U-U correlation function (SOAR Model, *c = 0.1 rad*).

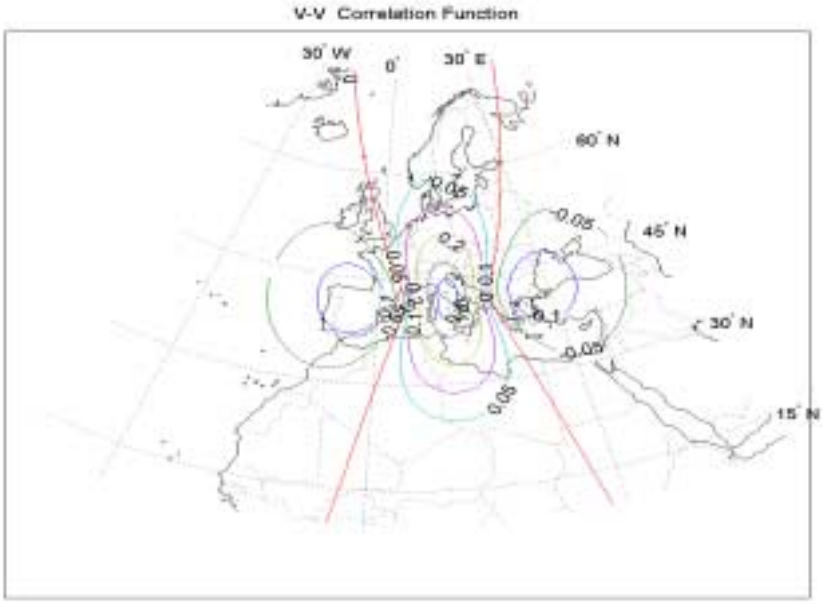

Figure 2.5 V-V correlation function (SOAR Model, *c = 0.1 rad*).

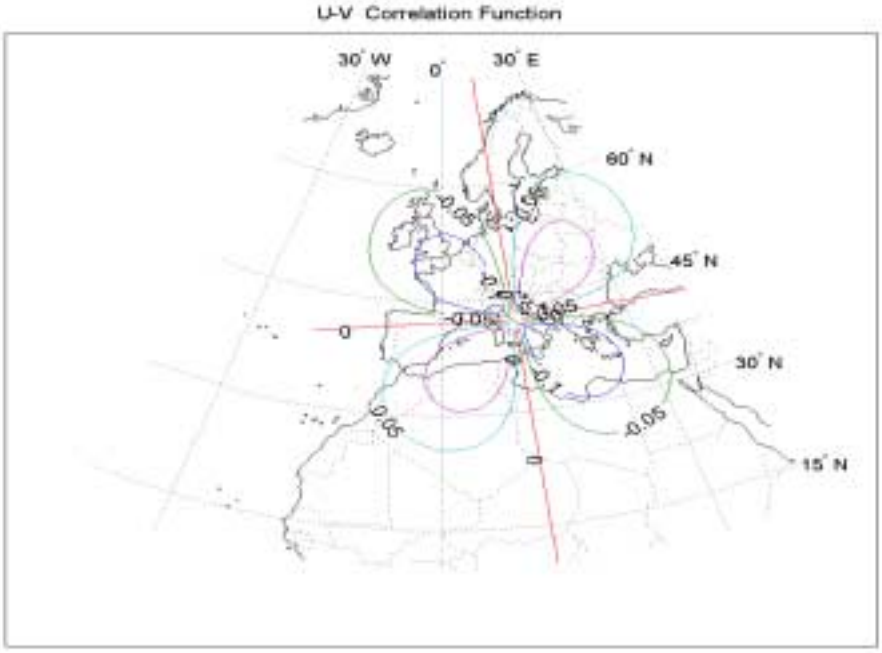

Figure 2.6 U-V correlation function (SOAR Model, *c = 0.1 rad*).

The vertical part of the correlation function has been modeled after Bergman (1979):

$$\chi^{\boldsymbol{uu}}(\boldsymbol{p}_i,\boldsymbol{p}_j) = (1+k_p \log^2(p_i/p_j))^{-1} \tag{2.68}$$

which, for the other correlations, gives:

$$\chi^{\boldsymbol{vv}}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \chi^{\boldsymbol{uv}}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \chi^{\boldsymbol{vu}}(\boldsymbol{p}_i,\boldsymbol{p}_j) = (1+k_p \log^2(p_i/p_j))^{-1} \tag{2.69}$$

$$\chi^{\boldsymbol{uT}}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \chi^{\boldsymbol{vT}}(\boldsymbol{p}_i,\boldsymbol{p}_j) = 2k_p^{1/2} \log(\boldsymbol{p}_i,\boldsymbol{p}_j) \; (1+k_p \log^2(p_i/p_j))^{-2} \tag{2.70}$$

$$\chi^{\boldsymbol{Tu}}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \chi^{\boldsymbol{Tv}}(\boldsymbol{p}_i,\boldsymbol{p}_j) = -2k_p^{1/2} \log(\boldsymbol{p}_i,\boldsymbol{p}_j) \; (1+k_p \log^2(p_i/p_j))^{-2} \tag{2.71}$$

$$\chi^{\boldsymbol{TT}}(\boldsymbol{p}_i,\boldsymbol{p}_j) = (1- 4k_p \log^2(\boldsymbol{p}_i,\boldsymbol{p}_j)\, \chi^{\boldsymbol{uu}}(\boldsymbol{p}_i,\boldsymbol{p}_j)) \; (1+k_p \log^2(p_i/p_j))^{-2} \tag{2.72}$$

Sketches of the functions are given below for $k_p=5$, in Fig. 2.7 through 2.10.

The vertical correlation parameter k_p will be determined through a statististical analysis of the observed minus forecast increments, as will be shown in Chapter 3.

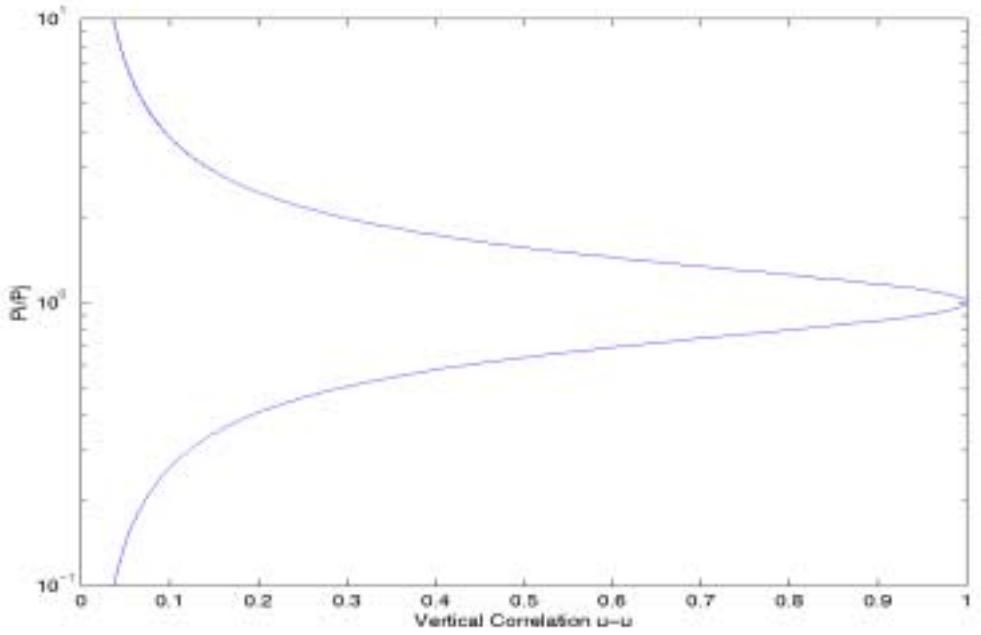

Figure 2.7 U-U Vertical Correlation function ($k_p = 5$)

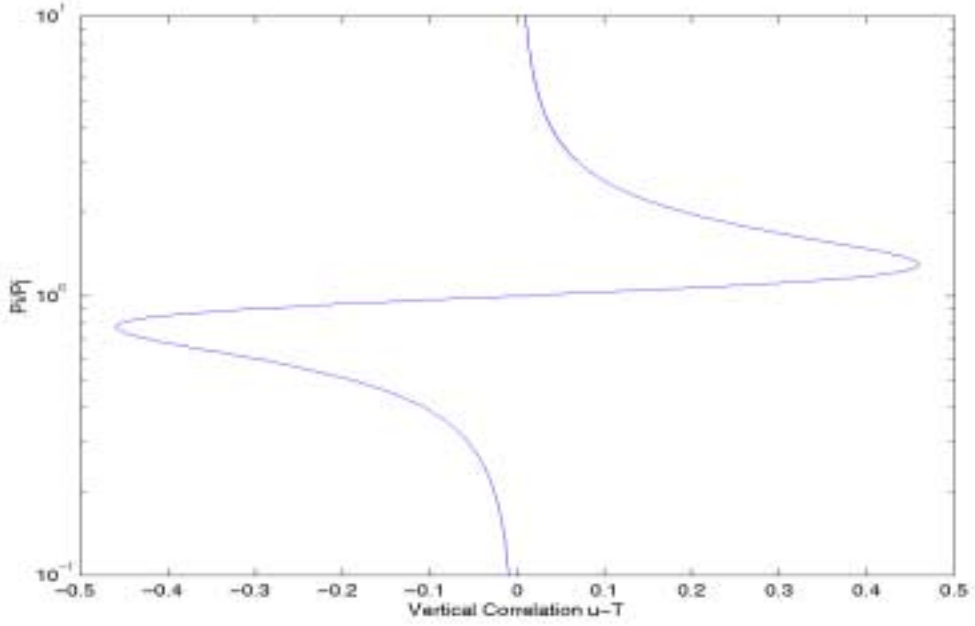

Figure 2.8 U-T Vertical Correlation function ($k_p = 5$)

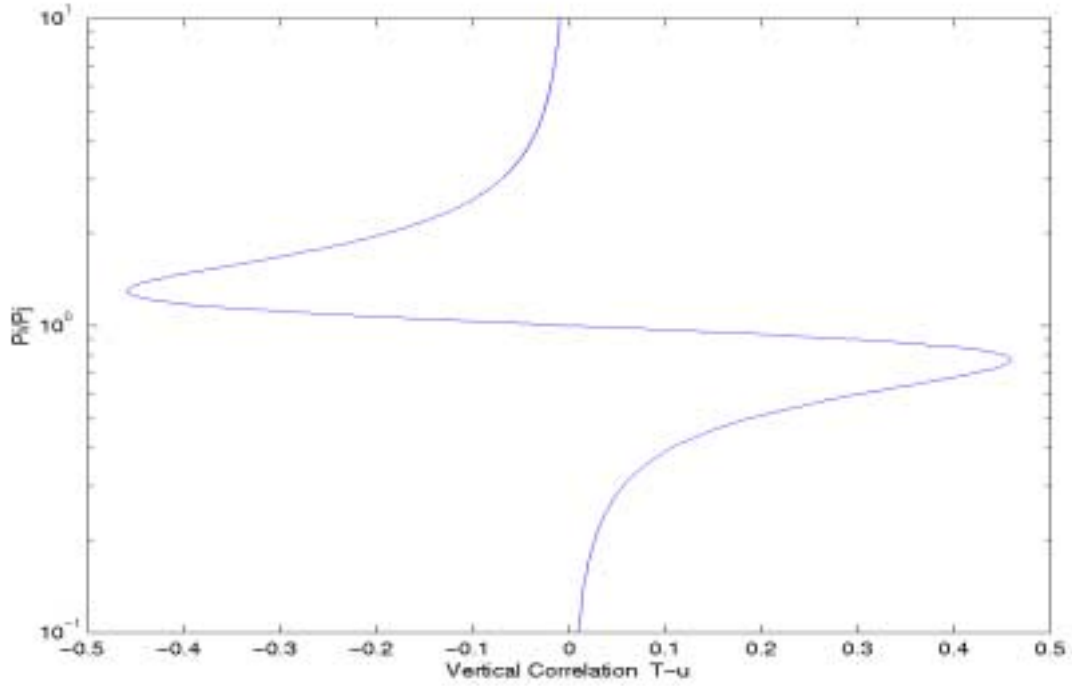

Figure 2.9 T-U Vertical Correlation function (k_p = *5*)

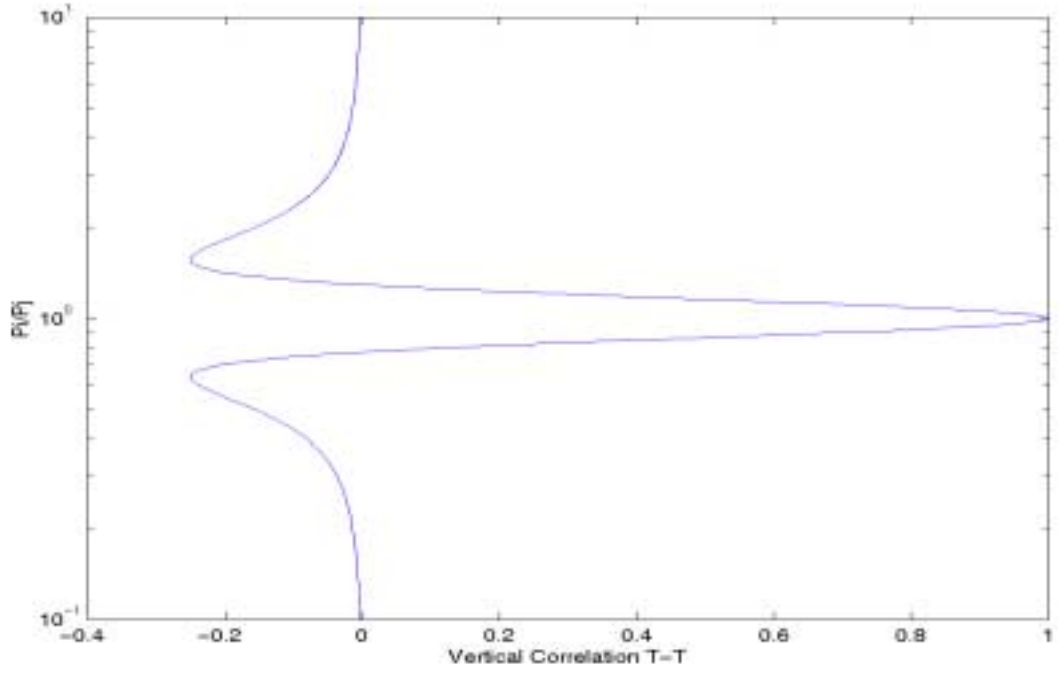

Figure 2.10 T-T Vertical Correlation function (k_p = *5*)

It is also useful to derive the mutual correlations of temperature and wind field with the geopotential. In order to do this, elimination of u,v from (2.26-2.29) leads to:

$$\partial \boldsymbol{T}/\partial\varphi = -1/(R_d)\ \partial^2 \boldsymbol{\Phi}/\partial\varphi\partial ln(\boldsymbol{p}) \tag{2.73}$$

$$\partial \boldsymbol{T}/\partial\lambda = -1/(R_d)\ \partial^2 \boldsymbol{\Phi}/\partial\lambda\partial ln(\boldsymbol{p}) \tag{2.74}$$

From these relations, assuming that the geopotential autocovariance can modeled as:

$$Cov(\boldsymbol{\Phi}_i,\ \boldsymbol{\Phi}_j) = \sigma_{\boldsymbol{\Phi}}^2\ R(\tau)\ \chi^{\boldsymbol{\Phi\Phi}}(\boldsymbol{p}_i,\boldsymbol{p}_j) \tag{2.75}$$

we can derive the correlations:

$$\partial/\partial\varphi_{\boldsymbol{i}}\ Cov(\boldsymbol{T}_i,\boldsymbol{\Phi}_j) = -\sigma_{\boldsymbol{\Phi}}^2 R_d^{-1}\ dR(\tau)/d\tau\ \partial\tau/\partial\varphi_i \partial/\partial ln(\boldsymbol{p}_i)\ \chi^{\boldsymbol{\Phi\Phi}}(\boldsymbol{p}_i,\boldsymbol{p}_j) \tag{2.76}$$

$$\partial/\partial\lambda_{\boldsymbol{i}} Cov(\boldsymbol{\Phi}_i,\boldsymbol{T}_j) = -\sigma_{\boldsymbol{\Phi}}^2 R_d^{-1}\ dR(\tau)/d\tau\ \partial\tau/\partial\lambda_j \partial/\partial ln(\boldsymbol{p}_j)\ \chi^{\boldsymbol{\Phi\Phi}}(\boldsymbol{p}_i,\boldsymbol{p}_j) \tag{2.77}$$

$$\partial^2/\partial\lambda_{\boldsymbol{i}}\partial\varphi_{\boldsymbol{i}}\ Cov(\boldsymbol{T}_i,\boldsymbol{T}_j) = -\sigma_{\boldsymbol{\Phi}}^2 R_d^{-2}\ dR(\tau)/d\tau\ \partial^2\tau/\partial\lambda_{\boldsymbol{i}}\partial\varphi_i\ \partial^2/\partial ln(\boldsymbol{p}_i)\partial ln(\boldsymbol{p}_j)\chi^{\boldsymbol{\Phi\Phi}}(\boldsymbol{p}_i,\boldsymbol{p}_j) \tag{2.78}$$

Separating the vertical and isobaric components of the correlations, we find:

$$\chi^{\boldsymbol{T\Phi}}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \partial/\partial ln\boldsymbol{p}_i \chi^{\boldsymbol{\Phi\Phi}}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \chi^{\boldsymbol{Tu}}(\boldsymbol{p}_i,\boldsymbol{p}_j) \tag{2.79}$$

$$\chi^{\boldsymbol{\Phi T}}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \partial/\partial ln\boldsymbol{p}_j \chi^{\boldsymbol{\Phi\Phi}}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \chi^{\boldsymbol{uT}}(\boldsymbol{p}_i,\boldsymbol{p}_j) \tag{2.80}$$

$$\chi^{\boldsymbol{TT}}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \partial^2/(\partial ln\boldsymbol{p}_i\partial ln\boldsymbol{p}_j)\,\chi^{\boldsymbol{\Phi\Phi}}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \chi^{\boldsymbol{uu}}(\boldsymbol{p}_i,\boldsymbol{p}_j) \tag{2.81}$$

while the isobaric components are given by:

$$\sigma_{\boldsymbol{T}}\sigma_{\boldsymbol{\Phi}}\, R^{\boldsymbol{\Phi T}}(\boldsymbol{r}_i,\boldsymbol{r}_j) = -\sigma_{\boldsymbol{\Phi}}^2\ R_d^{-1}\ dR(\tau)/d\tau\ \partial\tau/\partial\lambda_j \tag{2.82}$$

$$\sigma_{\boldsymbol{T}}\sigma_{\boldsymbol{\Phi}}\, R^{\boldsymbol{T\Phi}}(\boldsymbol{r}_i,\boldsymbol{r}_j) = -\sigma_{\boldsymbol{\Phi}}^2\ R_d^{-1}\ dR(\tau)/d\tau\ \partial\tau/\partial\varphi_i \tag{2.83}$$

$$\sigma_{\boldsymbol{T}}^2\ R^{\boldsymbol{TT}}(\boldsymbol{r}_i,\boldsymbol{r}_j) = \sigma_{\boldsymbol{\Phi}}^2\ R_d^{-2}\partial^2\tau/\partial\lambda_{\boldsymbol{i}}\partial\varphi_i\ dR(\tau)/d\tau \tag{2.84}$$

From these expressions we are also able to derive the geostrophically constrained autocovariances:

$$Lim_{(\tau \to 0)}\, Cov(\boldsymbol{T}_i, \boldsymbol{T}_j) = \sigma_{\boldsymbol{T}}^2 = R_d^{-2}\, Lim_{(\tau \to 0)}\, Cov(\boldsymbol{\Phi}_i, \boldsymbol{\Phi}_j) = R_d^{-2}\, \sigma_{\boldsymbol{\Phi}}^2 \tag{2.85}$$

In analogous fashion, starting from (2.25-2.26) the cross correlations between the wind components u,v and the geopotential Φ can be derived. The results of the computations are:

$$\chi^{u\Phi}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \chi^{\Phi u}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \chi^{\Phi\Phi}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \chi^{uu}(\boldsymbol{p}_i,\boldsymbol{p}_j) \tag{2.86}$$

$$\chi^{v\Phi}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \chi^{\Phi v}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \chi^{\Phi\Phi}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \chi^{uu}(\boldsymbol{p}_i,\boldsymbol{p}_j) \tag{2.87}$$

while for the isobaric cross correlations we find:

$$\sigma_u \sigma_{\Phi}\, R^{u\Phi}(\boldsymbol{r}_i,\boldsymbol{r}_j) = -\sigma_{\Phi}^2\, k/\sin\varphi_i\, dR(\tau)/d\tau\, \partial\tau/\partial\varphi_i \tag{2.88}$$

$$\sigma_{\Phi} \sigma_u\, R^{\Phi u}(\boldsymbol{r}_i,\boldsymbol{r}_j) = -\sigma_{\Phi}^2\, k/\sin\varphi_j\, dR(\tau)/d\tau\, \partial\tau/\partial\varphi_j \tag{2.89}$$

$$\sigma_v \sigma_{\Phi}\, R^{v\Phi}(\boldsymbol{r}_i,\boldsymbol{r}_j) = \sigma_{\Phi}^2\, k/(\sin\varphi_i \cos\varphi_i)\, dR(\tau)/d\tau\, \partial\tau/\partial\lambda_i \tag{2.90}$$

$$\sigma_{\Phi} \sigma_v\, R^{\Phi v}(\boldsymbol{r}_i,\boldsymbol{r}_j) = -\sigma_{\Phi}^2\, k/(\sin\varphi_j \cos\varphi_j)\, dR(\tau)/d\tau\, \partial\tau/\partial\lambda_j \tag{2.91}$$

2. Surface Analysis Covariance Model

The surface analysis covariances used in the objective analyses of the surface pressure (SP), mean sea level pressure (MSLP) and 10-meter wind fields are a simplified version of the models used in the upper air analysis.

In this case only the geostrophic constraints (2.26-2.27) apply. The computations are similar to the ones described above and lead to analogous results, with the notable exception that, assuming that the pressure autocovariance is modeled as:

$$Cov(\boldsymbol{p}_i,\boldsymbol{p}_j) = \sigma_p^2\, R(\tau) \tag{2.92}$$

Then, the geostrophically constrained wind variances are given by:

$$Lim_{(\tau \to 0)}\, Cov(\boldsymbol{u}_i, \boldsymbol{u}_j) = \sigma_u^2 = Lim_{(\tau \to 0)}\, Cov(\boldsymbol{v}_i, \boldsymbol{v}_j) = \sigma_v^2 = -(\sigma_p \boldsymbol{k}/\rho sin\varphi_i)^2 L \qquad (2.93)$$

where L is given by (2.47). For the SOAR model of (2.24) the wind component autocovariances reduce to:

$$\sigma_u^2 = \sigma_v^2 = (\sigma_p kc/\rho sin\varphi_i)^2 \qquad (2.94)$$

It is clear that this correlation model is really applicable only under the provision that the geostrophic components of the observed surface winds have been extracted. This requires the use of an appropriate boundary layer model. Work in this direction is under way, making use of the Planetary Boundary Layer model developed by R.A. Brown and coworkers (Brown and Levy, 1986).

3. Anisotropy and Flow Dependency of Covariance Functions

An implicit assumption of 3D-Var algorithms (and, in general, of intermittent assimilation schemes) is the stationary character of the background error covariances. This simplification, together with those of isotropy and homogeneity, is widely used in operational settings due to the reason that taking into account the temporal evolution of the background covariances would greatly increase the complexity of the objective analysis algorithm and the burden placed on the computing resources. However, it is very well known (Daley, 1991; Otte et al., 2001), that the observed-minus-background correlation patterns of weather systems show considerable anisotropy (the SW-NE tilt of upper level trough axis, for instance), which make the isotropy and stationarity approximations serious shortcomings of data assimilation systems which do not take the time evolution of the covariance matrices into account.

The problem has been tackled in a variety of ways. In the context of variational assimilation there are two main approaches:

1. the *4D-Var* scheme, where the background covariance matrix is evolved implicitly by the dynamics of the tangent linear model (and its adjoint; see Rabier and Courtier (1992) for details);
2. the *Kalman-Bucy* filter, wich explicitly evolves the background covariance matrix, and the many approximations which have been proposed in order to make this method computationally tractable ("Ensemble Kalman filter", Evensen,1994; "Reduced-Rank Kalman filter", Fisher, 1998; "The Cycling Representer algorithm", Xu and Daley,2000);

Only the 4D-Var approach has been implemented in operational environments even though many approximations are used in practice (ECMWF: Rabier et al., 2000). The Kalman filter method is under active study, but it does not seem feasible to be implemented in an operational setting with the current generation of computers. This is due to the fact that the method requires an explicit computation of the analysis error covariance matrix and its evolution in time and this matrix, for current numerical weather prediction models, is of the order of $10^7 \text{x} 10^7$.

Due to the unavailability at CNMCA of the human and computer resources necessary for tackling the problem of the temporal extension of 3D-Var, possible ways to circumvent the main shortcomings of the algorithm were investigated. The approach we have chosen was pioneered by Benjamin (1989) and it has recently been actively investigated by many researchers (Miller and Benjamin, 1992; Dévényi and Schlatter, 1994; Riishojgaard, 1998; Otte et al., 2000).

In this method the isotropic, stationary covariance model described in the preceding paragraphs, is modified through multiplication by an anisotropic, flow-dependent term that is a function of the background field potential temperature. For example, the temperature autocovariance function (2.29) is modified as follows:

$$Cov(\boldsymbol{T}_i, \boldsymbol{T}_j) = \sigma_{\boldsymbol{T}}^2 \, R(\tau) \, \chi^{\boldsymbol{TT}}(\boldsymbol{p}_i, \boldsymbol{p}_j) \; \nu(\,||\theta^b(\boldsymbol{r}_i, \boldsymbol{p}_i) - \theta^b(\boldsymbol{r}_i, \boldsymbol{p}_j)||\,) \qquad (2.95)$$

In the present implementation the anisotropic component ν is modeled by a simple SOAR function of the absolute difference of the background potential

temperatures at locations *i,j*. This is a computationally cheap way of obtaining flow-dependent covariances, and it relies heavily on the accuracy of the first guess potential temperature field in order to produce positive results. However, it might be argued that also in the Kalman filter approach the time evolution of the analysis error covariance matrix depends on the accuracy of the previous analysis step and of the model itself (actually, a linearized version of the model). We note in passing that the 3-D character of this formulation should also be helpful in the objective analysis of single level observations in the mass-wind analysis (SYNOP, SHIP, BUOY, AIREP, etc), where the use of the statistically derived vertical correlation functions (2.67-2.72) can be detrimental when it does not take into account the presence of sharp air-mass boundaries (i.e. boundary layers with sharp inversions, tropopause boundary).

An example of how the flow-dependent term impacts the objective analysis is now discussed. The main feature of the synoptic state of the atmosphere over the model domain on the 19th of June 2002, 00UTC, is the elongated upper level trough whose cyclonic part extends over northwestern Spain, the British Isles and Scandinavia (Fig. 2.11). In the lower troposphere this is mirrored by a sharp air mass boundary extending from Spain trough France, Germany and the southern portion of the Scandinavian Peninsula (Fig. 2.12). In this assimilation cycle a radiosonde near Paris reported a 500 hPa temperature 1.5°C higher than the forecast temperature. The way in which this observation increment is spatially interpolated in the ensuing objective analysis can be seen in Fig. 2.13. The air mass boundary present at 500 hPa (Fig. 2.14) clearly models the shape of the correlation function in an anisotropic and flow-dependent way.

Although these results look promising and intuitively appealing, the merits of the method must be evaluated in an objective way, through comparisons of statistical skill scores of the forecast fields derived from objective analysis performed both with and without the flow-dependent term. This work is in progress.

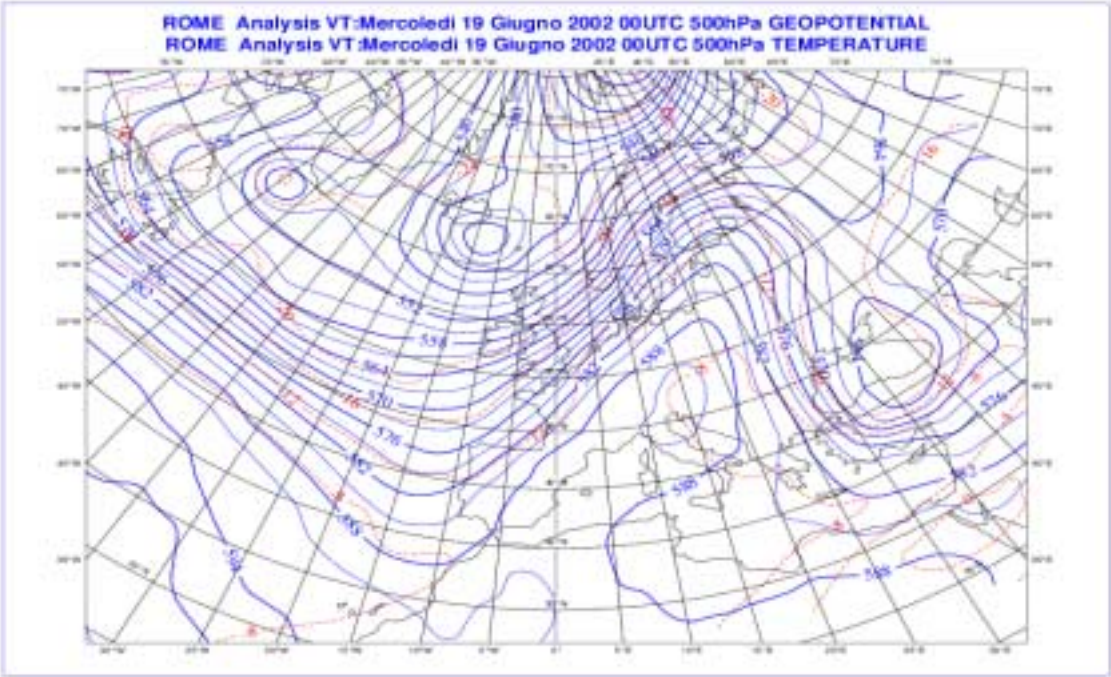

Figure 2.11 CNMCA 500 hPa Geopotential height and Temperature Analysis: June 19th 2002, 00UTC.

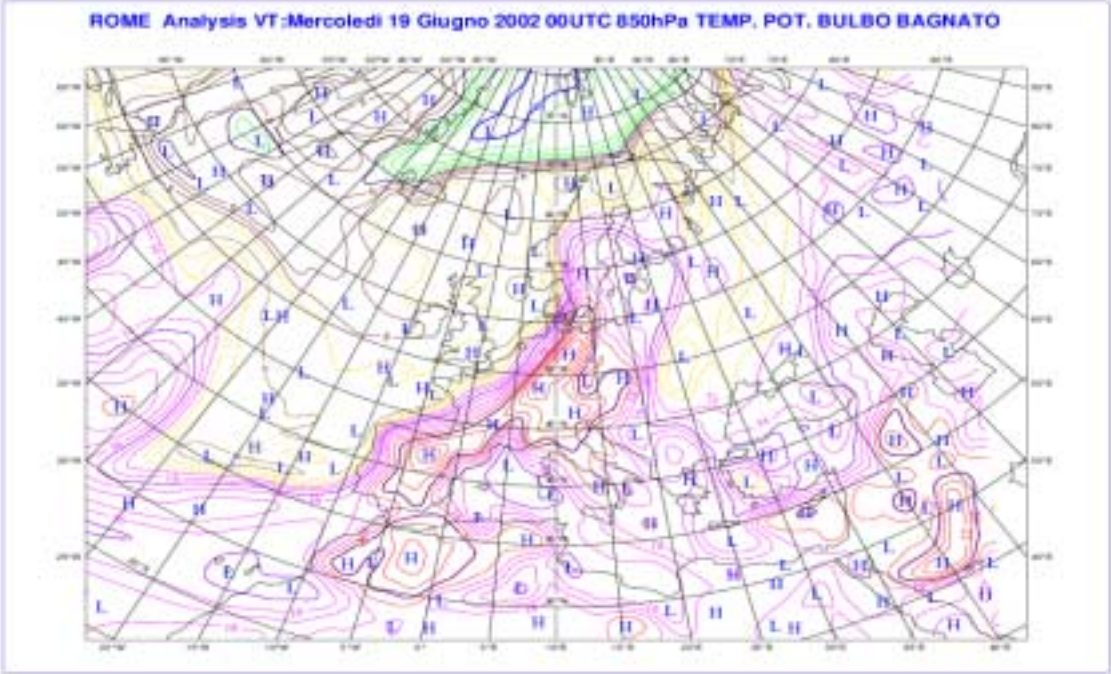

Figure 2.12 CNMCA 850 hPa Wet Bulb Potential Temperature Analysis: June 19th 2002, 00UTC.

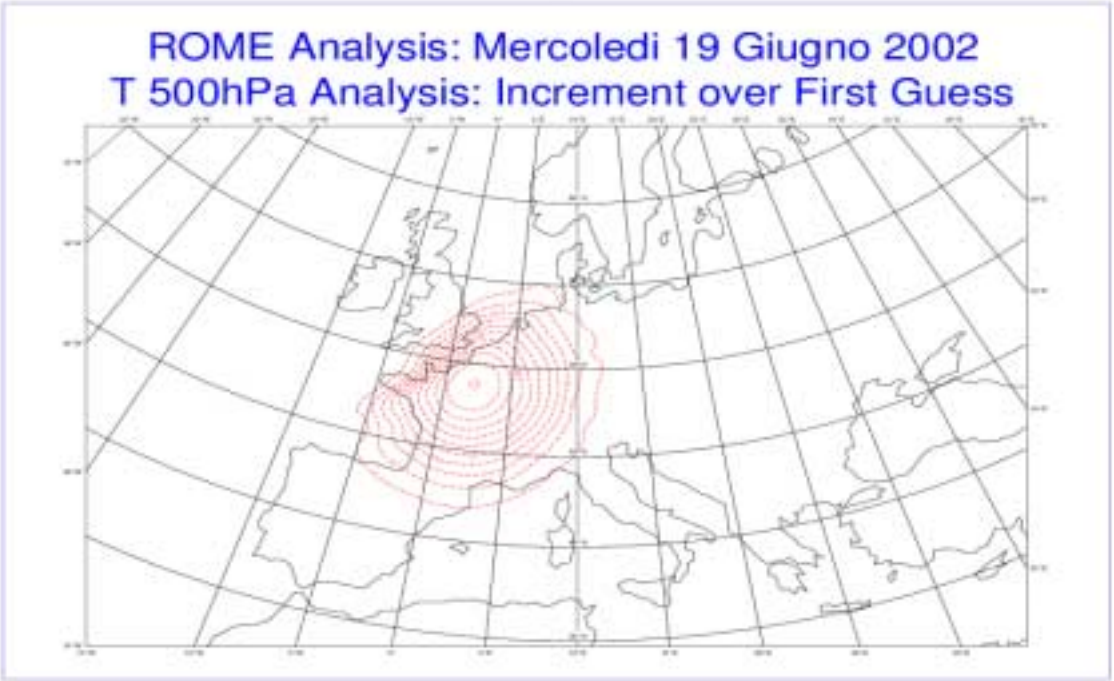

Figure 2.13 CNMCA 500 hPa Temperature Analysis Increment over First Guess: June 19th 2002, 00UTC.

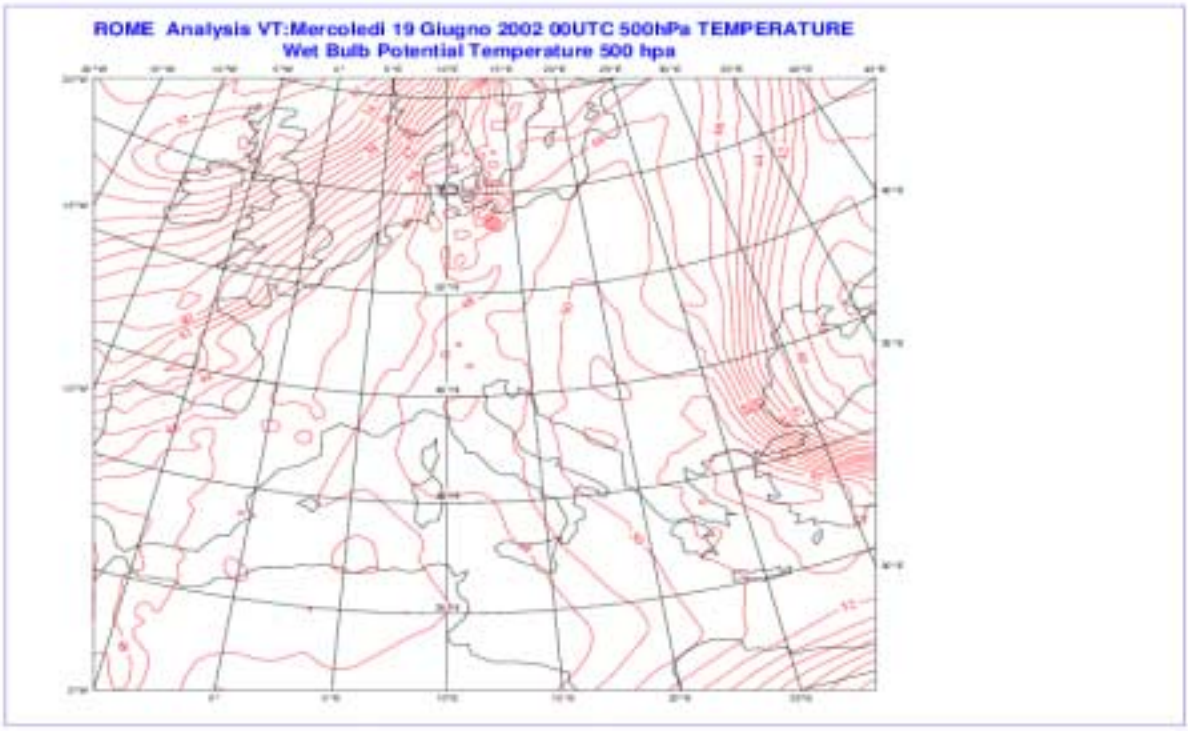

Figure 2.14 CNMCA 500 hPa Potential Temperature Analysis: June 19th 2002, 00UTC.

THIS PAGE INTENTIONALLY LEFT BLANK

III. BACKGROUND AND OBSERVATION ERRORS STATISTICS

In order to maximize the amount of information extracted from the observations and, at the same time, reject spurious noise (i.e., information related to spatial scales which the numerical model and, consequently, the objective analysis algorithm, cannot resolve), it is vital that the background ($\boldsymbol{P}_b$) and the observation ($\boldsymbol{R}$) error covariance matrices be specified as accurately as possible.

Recalling their definitions:

$$\boldsymbol{P}_b = < \boldsymbol{e}_b \boldsymbol{e}_b^T > \tag{3.1}$$

$$\boldsymbol{R} = < \boldsymbol{e}_o \boldsymbol{e}_o^T > \tag{3.2}$$

and the fact that both errors $\boldsymbol{e}_b$ and $\boldsymbol{e}_o$ refer to unknown, unbiased, "true" values, it is clear that we cannot compute these quantities from first principles. What is done in practice is to estimate these covariances by a statistical procedure: we are assuming that these errors are stationary in time and uniform over our spatial domain in order to derive their "climatology". From this and the underlying assumptions that the errors are normally distributed and unbiased, we are then able to compute the second moments (variances and covariances) of their probability density functions (pdf).

In practice this calculation can be done in at least three different ways:

1. The "*Observation method*" (Rutherford, 1972; Hollingsworth and Lönnberg, 1986). In this procedure observation-minus-first guess differences are collected for a period of time over a network of homogeneous and uncorrelated observing stations. From this data, which can be further stratified by pressure level and time of the year, the spatial statistics of the covariances can be computed, together with the perceived background and observation error variances;

2. The "*NMC method*" (Parrish and Derber, 1992). The main assumption of the method is that the statistical structure of the background correlations does not change significantly over the first 48h of model integration. Then it is possible to derive the background covariances from a statistical analysis of the differences of the 48h and 24h forecast fields verifying at the same time. The method is suitable for global numerical prediction systems and also very practical and straightforward to set up. On the other hand the main hypothesis it rests upon is rather difficult to justify;

3. The "*Analysis-Ensemble method*" (Fisher, 2001). This method is the one currently used at ECMWF. An ensemble of different analyses is realized by randomly perturbing the initial observations within the assumed observation error. These different analyses are then integrated in time till the next objective analysis, where the process is repeated. After a few days the differences in the first guess fields should be representative of the underlying background error statistics.

The method chosen to specify the parameters involved in the modeling of the background and observation error matrices is the *Observation* method. The reason for this is that it is a direct and theoretically sound way of deriving the background spatial correlations and, as a bonus, it is also capable of providing an estimate of the background and observation variances, whose relative magnitude is perhaps the most fundamental quantity to be specified in every objective analysis algorithm.

The main steps in the computations and the more relevant results will be described in the following paragraphs. For more details, see Vocino (2002). Also an account of the observations currently used and their assumed error statistics will be given.

A. BACKGROUND ERROR STATISTICS

In order to derive the necessary statistics on the background error, the "*Observation method*" has been used. The entire network of land radiosonde stations

present in the analysis domain (165) has been exploited in the computations: the data refer to the 00Z and 12Z synoptic times over a three month period starting from the 13th of March 2002. The temperature 6-h observation increments (denoted as O_k-B_k) were collected for the standard pressure levels (1000 hPa, 925 hPa, 850 hPa, 700 hPa, 500 hPa, 400 hPa, 300 hPa, 200 hPa, 100 hPa, 50 hPa).

After removing the bias for each station, the following estimate of observation increments correlations was computed for each pair of stations *k,l*:

$$R_{kl} = <(O_k-B_k)\ (O_l-B_l)> /(<(O_k-B_k)^2>^{1/2}< (O_l-B_l)^2>^{1/2} \quad (3.1)$$

Due to the separability assumption, we can study the quasi-horizontal (isobaric) correlations independently from the vertical correlations.

In order to study the isotropic component of the isobaric correlations $R_{kl,}$ they have been partitioned into intervals of 0.01 rad (≈200 Km) of their mutual great circle distance. In each of these intervals the Fisher z-transform of the empirical correlation coefficients has been performed:

$$Z = 1/2\log((1+R)/(1-R)) \quad (3.2)$$

This has been done in order to preserve the assumed normal distribution of the correlation coefficients around their mean value, and it has proved beneficial in the successive fit of the correlation models to the empirical data (Fig 3.1).

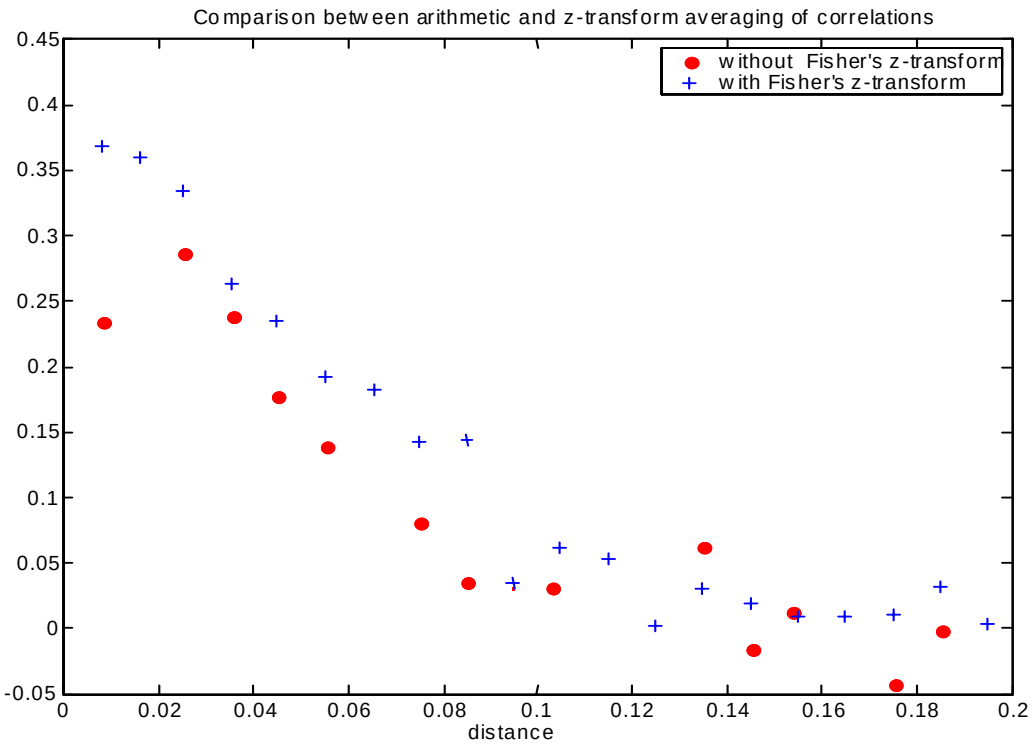

Figure 3.1 Example of correlation fit with arithmetical and Fisher z-transform mean.

The model functions fitted to the empirical correlation coefficients have been the Negative Squared Exponential (NSE) function:

$$\rho_B = exp(-0.5(r/L_c^2)) \tag{3.3}$$

and the Second Order Autoregressive (SOAR) function:

$$\rho_B = (1+r/L_c)exp-(r/L_c) \tag{3.4}$$

What we are trying to determine is the correlation length L_c, which can be defined for any correlation model as (Daley, 1991):

$$L_c \equiv (-2\rho/\nabla^2\rho)^{1/2}_{r=0} \tag{3.5}$$

where the Laplacian operator ∇^2, due to the isotropy assumption, reduces to:

$$\nabla^2 = 1/r \; d/dr(rd/dr) \tag{3.6}$$

Sample results of the fits are given below (distances are in radians units of Earth's radius).

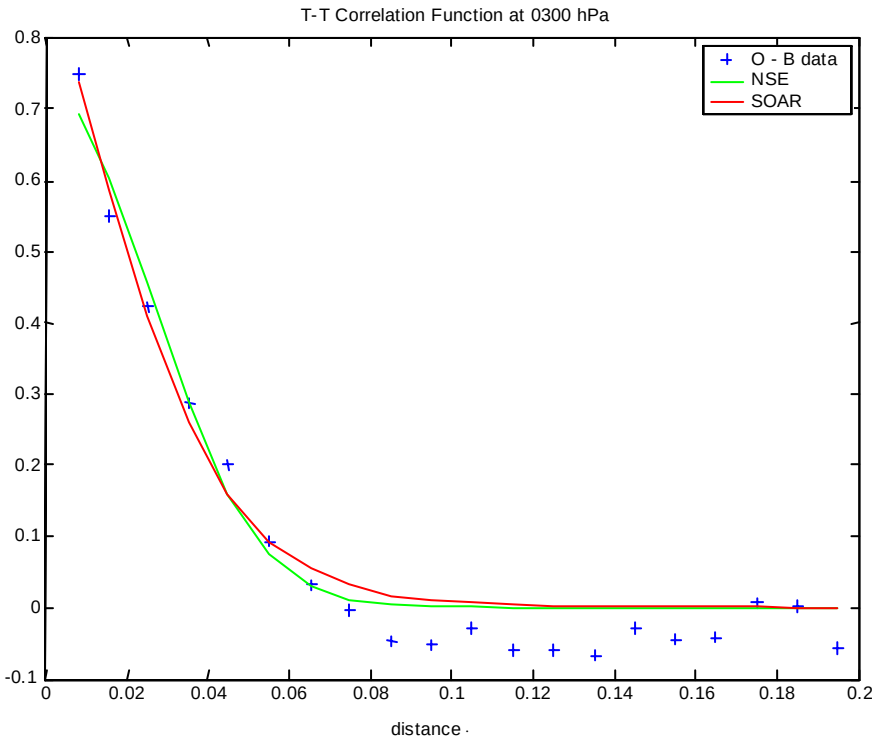

Figure 3.2 Correlation fit for isobaric (300 hPa) Temperature Observation increments.

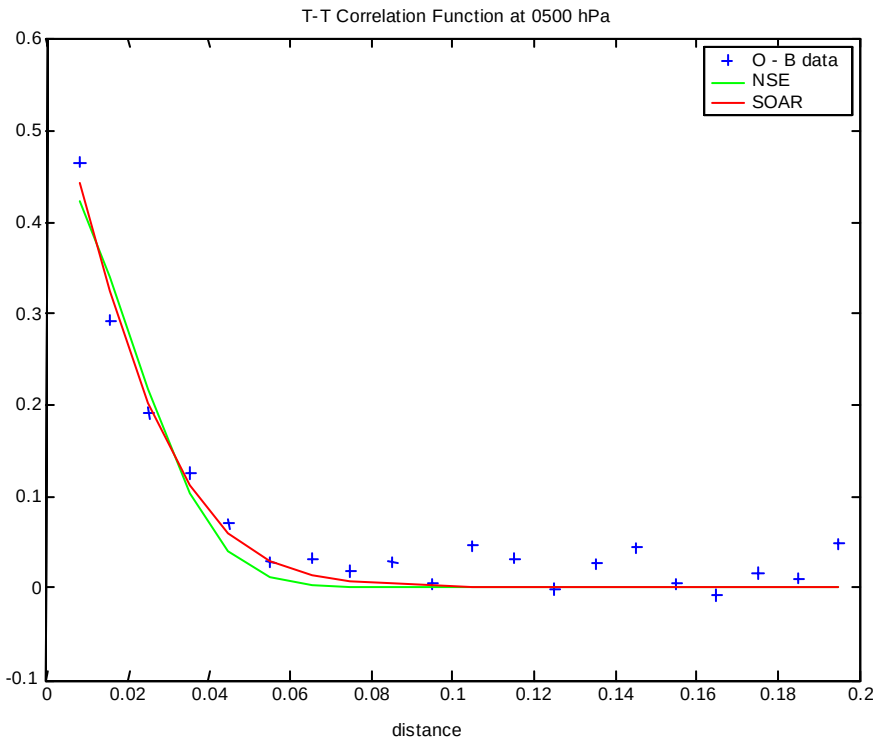

Fig. 3.3: Correlation fit for isobaric (500 hPa) Temperature Observation increments.

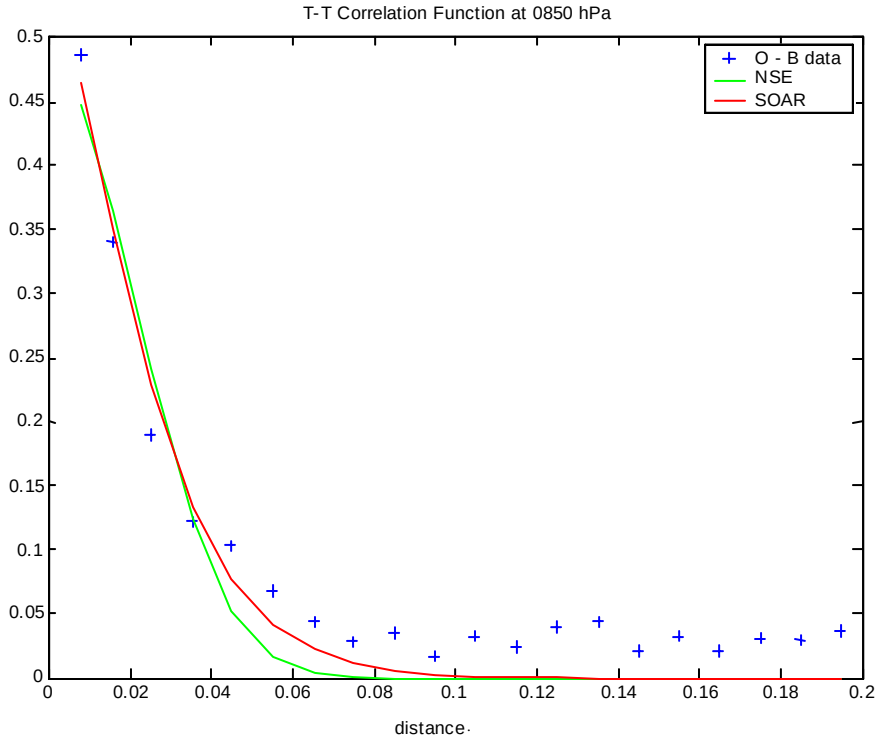

Fig. 3.4: Correlation fit for isobaric (850 hPa) Temperature Observation increments.

The main quantitative results are summarized in the following two tables:

Pressure Level (hPa)	L_c **(Km)**	**Fit rmse**
1000	713	0.0618
925	339	0.0321
850	426	0.0362
700	378	0.0244
500	409	0.0304
400	368	0.0250
300	517	0.0441
200	539	0.0214
100	893	0.0488
50	809	0.0421

Table 3.1 Correlation lengths and rmse of fit for Temperature Observation increments: NSE correlation model.

Pressure Level (hPa)	L_c (Km)	Fit rmse
1000	417	0.0560
925	201	0.0259
850	261	0.0291
700	221	0.0186
500	242	0.0251
400	220	0.0195
300	296	0.0443
200	318	0.0251
100	544	0.0329
50	482	0.0281

Table 3.2 Correlation lengths and rmse of fit for Temperature Observation increments: SOAR correlation model.

From inspection it is clear that the SOAR function gives a better fit to the empirical correlations at almost all levels: this is mainly due to the better agreement of the SOAR model function with the data at intermediate distances (0.04-0.08 radians), where the NSE function is seen to fall off too steeply. The correlation lengths have been plotted for ease of reference in Fig. 3.5.

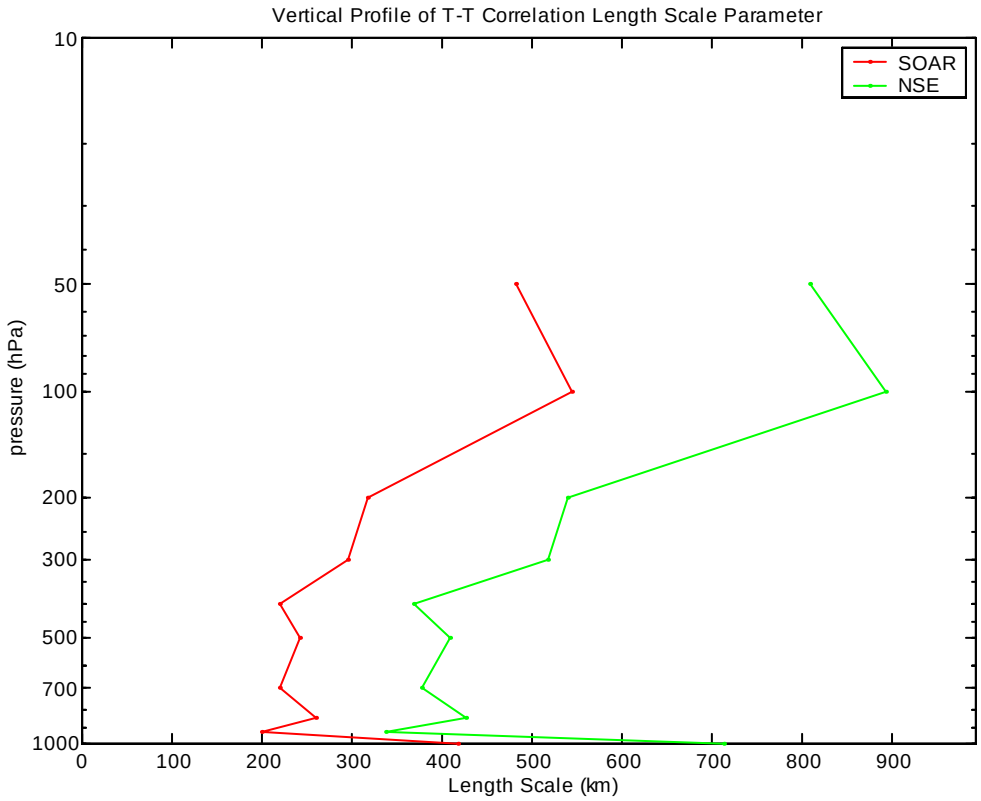

Figure 3.5 Vertical profiles of Correlation Lengths for Temperature Observation increments.

Apart from the anomalously high value at 1000 hPa (whose cause is under investigation), the correlation lengths appear to have a plausible height profile. They are approximately constant over much of the troposphere and increase considerably in the stratosphere. The lower correlation length values shown by the SOAR model agree with the more gradual incline of the function at large distances, where the linear term becomes dominant.

For the experimental runs of the objective analysis procedure, the SOAR model has been selected, with a constant tropospheric correlation length of L_c = 250 Km and a constant stratospheric (i.e. $p_{lev} \geq 250$ hPa) correlation length of L_c = 400 Km.

Another appealing feature of the "*Observation method*" is the possibility of deriving objective estimates of the background and observation errors' variances, which arguably

are the most important parameters to be specified in an objective assimilation algorithm. How this is done follows from the definition of the correlation coefficient R_{kl} (3.1): assuming homogeneity of the errors and uncorrelated observation errors (as is the case for independent radiosonde measurements), it can be shown (Daley, 1991) that:

$$R_z \equiv lim_{r \to 0} R_{kl}(r) = E_B^2/(E_B^2+E_O^2) \tag{3.7}$$

where E_B^2, E_O^2 are the background field and observation variances respectively. This is what can intuitively be expected, since the observations are mutually uncorrelated: for distances r close, but not equal to 0, the correlation can only be explained by the background term contribution, so that extrapolating to zero distance gives the relative weight of the background term with respective to the total (background + observation) variance.

On the other hand the total variance of the observation increments can be easily computed from:

$$E_B^2+E_O^2 = 1/K \, \Sigma_k <(O_k-B_k)^2> \tag{3.8}$$

so that, using (3.7-3.8) each variance can be calculated. The results of these computations have been summarized in Tables 3.3 and 3.4.

Pressure Level (hPa)	E_b(°C)	E_o(°C)	E_t(°C)
1000	1.52	2.19	2.67
925	1.74	1.19	2.36
850	1.26	1.31	1.82
700	1.24	1.10	1.66
500	1.46	1.58	2.16
400	1.56	1.64	2.26
300	2.24	1.37	2.63
200	2.72	0.94	2.88
100	1.15	1.16	1.63
50	1.25	1.32	1.82

Table 3.3 Background error, Observation error and Total perceived error for Temperature Observation increments: NSE correlation model.

Pressure Level (hPa)	E_b(°C)	E_o(°C)	E_t(°C)
1000	1.62	2.12	2.67
925	1.86	1.45	2.36
850	1.33	1.24	1.82
700	1.32	1.00	1.66
500	1.56	1.49	2.16
400	1.66	1.54	2.26
300	2.39	1.08	2.63
200	2.87	0.21	2.88
100	1.20	1.10	1.63
50	1.32	1.25	1.82

Table 3.4 Background error, Observation error and Total perceived error for Temperature Observation increments: SOAR correlation model.

Also these values have been plotted for ease of reference in Fig.3.6 and 3.7.

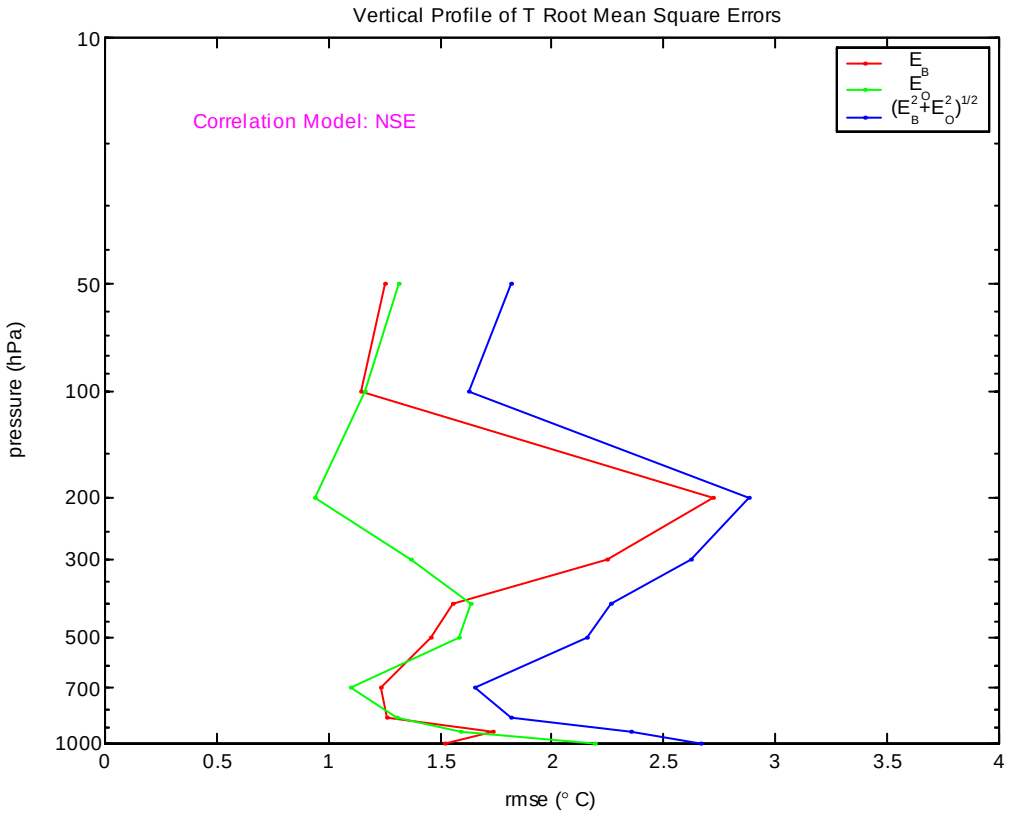

Figure 3.6 Vertical profile of Background error, Observation error and Total perceived error for Temperature Observation increments: NSE correlation model.

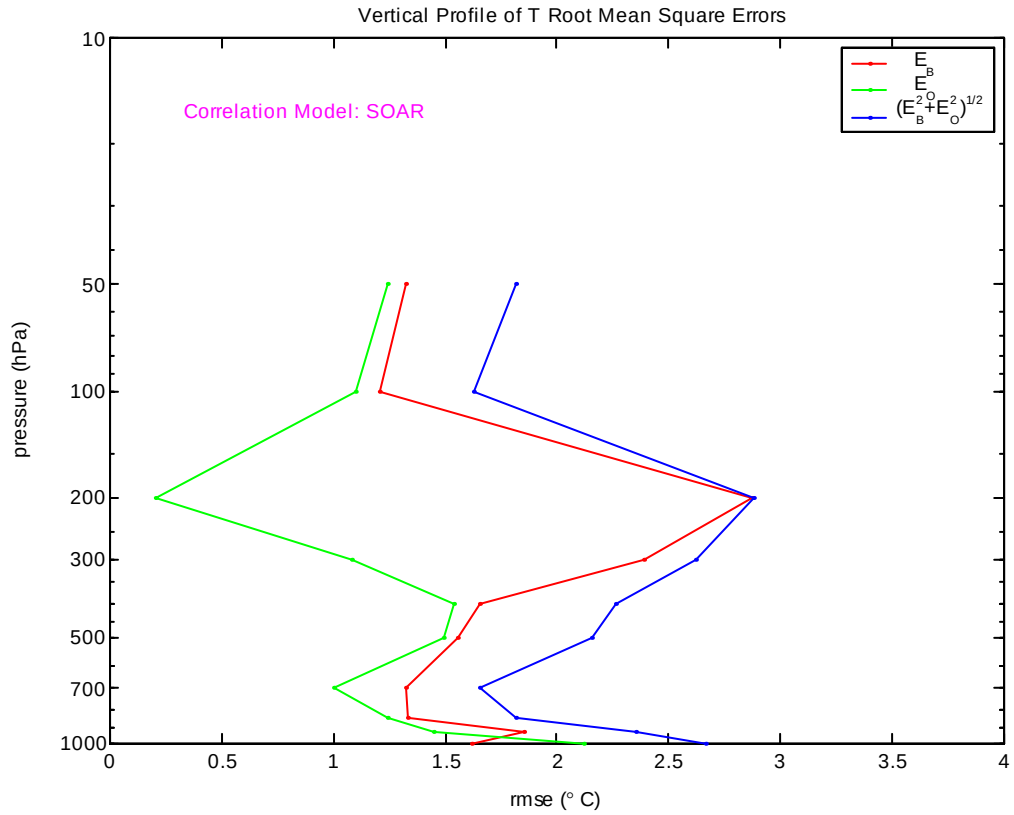

Figure 3.7 Vertical profile of Background error, Observation error and Total perceived error for Temperature Observation increments: SOAR correlation model.

From the plots it is clear that, with the exception of the 200 hPa level, the background and observed root mean square errors are fairly constant with height and of comparable magnitude, in the range of 1.2-1.5°C: these values are compatible with the expected accuracy of radiosonde observations and 6-h forecast fields.

The statistical calculation of the parameters in the vertical component has been performed in analogous fashion. Since, from inspection of 2.67-70, 2.78-80 and 2.85-86, it is clear that all vertical correlations can be expressed in term of $\chi^{\Phi\Phi}(p_i,p_j) = \chi^{uu}(p_i,p_j)$, the observed vertical correlations of the geopotential background increments have been fitted to the model function $(1+k_p log^2(p_i/p_j))^{-1}$ (Fig.3.8).

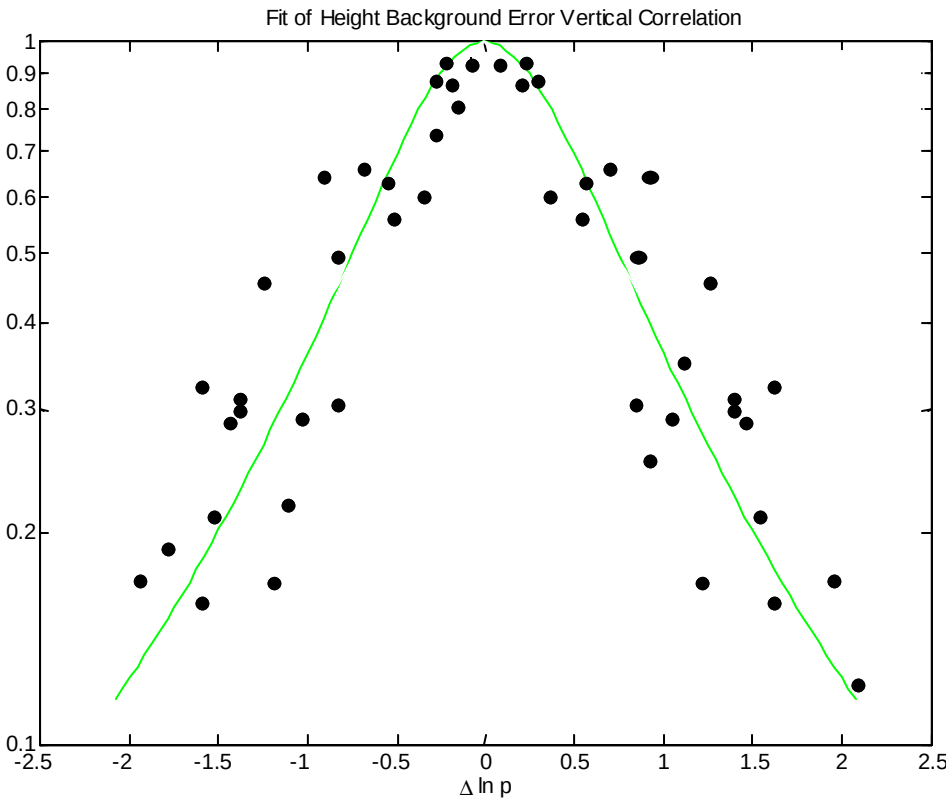

Fig. 3.8: Fit of vertical profile of geopotential height background minus observation increments: equation (2.68) correlation model.

The fit is satisfactory (*RMSE = 0.101*), giving for the fitting parameter *kp= 1.77,* which is the value used in the experimental trials of the objective analysis scheme.

B. OBSERVATION ERROR STATISTICS

Since the new objective analysis scheme is still in an experimental stage and it was felt it was important to validate the merits of the new algorithm and the revised background statistics and correlation functions, the observations used are the same ones used in the current operational Optimum Interpolation analysis. A brief account of their error statistics will be given below.

1. Radiosondes and Pibals

Even today, radiosonde reports (TEMP) are the main component of the global weather observing system in the Northern Hemisphere. On the main synoptic hours (00 and 12 UTC) about 165 reports are routinely analyzed (around 80 reports at 06 and 18 UTC synoptic hours). Mandatory and significant level observations of temperature, relative humidity and wind are assimilated. Observations at levels not coincident with isobaric analysis levels are vertically (linearly in *log(p)*) interpolated. From the results shown in the previous section, the assumed accuracy (RMSE) of the temperature observations have been set to 1.2ºC; no radiative corrections are employed. For relative humidity the assumed accuracy has been set to 5%. Perceived winds accuracy varies from $2.1 ms^{-1}$ in the lower levels to $3.5\ ms^{-1}$ in the stratosphere.

A major advantage of the new objective analysis is that it employs temperatures instead of geopotential heights, as the previous one did. This brings about two positive features: the temperature observation errors are much less vertically correlated than the geopotential errors and the background error correlations for temperature are sharper than the corresponding geopotential correlations. The end result of these two factors is that higher vertical resolution analysis is possible. Also, on a more technical level, making the assumption of vertically uncorrelated temperature observation errors drastically improves the condition number of the analysis equations, thus greatly reducing the computation time required for the minimization of the cost function. Pilot weather balloons (Pibals) wind reports are also routinely analyzed, with expected observation errors somewhat greater than those assigned to radiosonde reports.

2. Aircraft Based Observations

Aircraft based temperature and wind observations are routinely assimilated, after they have undergone the preprocessing step of the analysis cycle, where consistency checks on the moving platform reported position and climatological gross limit checks are performed.

Automatic observations (AMDAR, Aircraft Meteorological Data Relay) are given greater weight (i.e., smaller observation RMSEs) over manual observations (AIREP, Aircraft Reports). Both are treated as uncorrelated, single level reports.

3. Atmospheric Motion Winds

Atmospheric motion winds are derived from tracking the movement of identified cloud structures from a sequence (i.e. three) of geostationary satellite images. These observations are not routinely analyzed in the current operational objective analysis, but their use is envisaged in the new data assimilation system. It is felt, however, that, due to the current uncertainties in the height assignment of the winds, research is still needed in order to design an observation operator H able to extract all the relevant information.

4. Conventional Surface Observations

Conventional surface observations, both over land (SYNOP) and over the sea (SHIP, BUOYS) are routinely analyzed. After preprocessing and superobbing the reports closer than the average grid spacing (≈55 Km), around 1200 observations are fed into objective analysis algorithm. For the upper air analysis surface pressure innovations are transformed into geopotential innovations of the closest analysis level and their impact on the analyzed variables (T,u,v) is evaluated through the use of the correlations 2.78-2.90.

Surface fields (Surface pressure, Mean Sea Level pressure, 10-m wind) are analyzed through a two-dimensional, multivariate version of the algorithm (see Section II.C.2). Winds are only assimilated over sea. At the moment geostrophy is not strictly enforced, using a geostrophic coupling parameter $\boldsymbol{\mu}$=0.7. However it is felt that a more accurate estimate of the geostrophic, balanced component of the observed winds is needed in order to make good use of the wind observations. Ways to accomplish this by taking into account the structure of the forecast marine boundary layer are currently being investigated.

The Mean Sea Level pressure (MSLP) is well defined over the sea but, over land, is more of a visual aid for forecasters than a real meteorological field. However it is felt to be important that the analyzed MSLP field closely draws to the accepted MSLP and surface wind observations. To accomplish this, it has been found necessary to objectively analyze the MSLP and surface wind departures. Reduction of surface pressure (SP) analysis to mean sea level did not give satisfactory results, mainly because of the differences between model temperatures and observing station temperatures.

5. Scatterometer Winds

Scatterometers are satellite-borne radars, which provide measurement of surface wind speed and direction over sea. There is ambiguity, however, in the wind directions, and four different wind directions are compatible with each measurement. This ambiguity can be resolved either by choosing the direction closer to the first guess or by more advanced methods that take into account the spatial coherence properties of the retrieved wind field.

At the moment work is under way to assimilate the wind field product generated by the Dutch Meteorological Institute (KNMI) from the QuikSCAT satellite observations (SeaWinds). This is "…a **near real-time** 100-km resolution QuikSCAT product, which includes inversion, Quality Control, and a 2D-Var ambiguity removal algorithm…" making it suitable for data assimilation purposes (for more information, see www.knmi.nl/onderzk/applied/scattmtr/quikscat/index.html).

THIS PAGE INTENTIONALLY LEFT BLANK

IV. VALIDATION AND PRELIMINARY RESULTS

There are many possible ways through which the quality of the objectively analyzed fields can be gauged:

1. A subjective, "synoptic" evaluation of the charts produced by the data assimilation cycle from a forecaster's perspective, assessing the adherence to the observations (especially those not used in the analysis scheme, such as satellite imagery) and the degree of enforcement of "physical" balance properties;

2. Use of internal diagnostics, such as the analysis error covariance matrix, which can be computed as

$$\boldsymbol{P}_a = (\boldsymbol{P}_b^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1} \tag{4.1}$$

   It is common to compute only the diagonal elements of the matrix, which give an estimate of the variances of the analyzed variables. Unfortunately eq. (4.1) strictly holds on condition that the background and observation error covariance matrices ($\boldsymbol{P}_b,\boldsymbol{R}$) have been correctly specified, which is not usually the case. If this is not true, then eq.(4.1) underestimates the real analysis error covariances

3. A statistical, "objective" verification through comparison of forecasts produced from the analyzed fields with other forecasts started from independent data assimilation cycles.

This later approach has been taken in this study. Details of its implementation will be given below, together with a discussion of the results so far obtained.

A. METHOD OF VERIFICATION

In order to assess the quality of the analysis fields in an objective manner, two parallel runs of the HRM model (see Section I.A.3) have been set up, running at 00UTC every day up to T+48h: one starting from the analysis fields of the new data assimilation cycle, the other from the analyzed ECMWF fields interpolated to the HRM model grid. Apart from the initial conditions, all the other features of the two model integrations are equal (boundary conditions, resolution, etc.). This should guarantee that any difference in the subsequent forecast fields should be traced back to differences in the initial conditions. The accuracy of the forecast fields is estimated through the use of two common scalar measures of gridded fields: the root mean square error (RMSE) and the anomaly correlation (AC).

Denoting by f_{ij} the grid values of the forecast field and a_{ij} the grid values of the verifying analysis, then the RMSE is given by (Wilkes, 1995):

$$RMSE = (1/(MN)\Sigma_{i=1,M}\Sigma_{j=1,N}(f_{ij} - a_{ij}))^{1/2} \qquad (4.1)$$

This skill score is what we might define as an absolute measure of the forecast's quality. However it is not able to distinguish if the errors are related to biases in the forecast fields or to the misplacement of significant weather patterns. In order to do this the anomaly correlation score is useful (Wilkes, 1995). If we define C_{ij} the climatological averages of the analyzed fields at each grid point, then:

$$AC = (\Sigma_{i=1,M}\Sigma_{j=1,N}(f_{ij} - C_{ij})\,(a_{ij} - C_{ij}))/((\,\Sigma_{i=1,M}\Sigma_{j=1,N}(f_{ij} - C_{ij})^2\,(\Sigma_{i=1,M}\Sigma_{j=1,N}(a_{ij} - C_{ij})^2)) \qquad (4.2)$$

From this expression it is clear that we are investigating the correlations between the anomalies with respect to climatology of the forecast and the analyzed fields. In this way we are highlighting the pattern similarities between the two fields, while giving less weight to their absolute values.

A final point to be made is that the verifying analyses are the ECMWF analyses interpolated on to the HRM grid. This choice has been made in order to test the quality of the forecast fields derived from the data assimilation cycle under their most unfavorable

conditions. The statistical errors so derived should then represent an upper limit to the expected forecast errors.

B. PRELIMINARY RESULTS

Some results from a statistical comparison of the model runs over 30 cases (15th June 2002 – 15th July 2002) are shown in Fig.4.1 through 4.8 for the mean sea level pressure, geopotential height and wind speed fields. From inspection of the charts, it is evident that:

1. Degradation in quality with respect to the ECMWF based model run is within acceptable limits (~ 0.7 hPa for MSLP, from 5 to 10 gpm for geopotential height, from 0.9 m/s at 850 hPa to 2.3 m/s at the jet level height for wind speed);

2. The anomaly correlation scores for both the MSLP and 500 hPa geopotential height field are very close: this is an indication that the location of weather system is correctly placed in the CNMCA analysis' derived forecast fields and that the main source of error is to be found in the diagnosed intensities;

3. The error evolution is smooth in time, indicating that the analysis fields are perturbing the first guess fields in a "balanced" and physically reasonable manner. The gradual decrease in the RMSE differences between the two model runs is to due to the steady increase of influence of the common ECMWF derived boundary conditions.

It should also be borne in mind the different spatial and vertical resolution of the ECMWF analyses (which are the results of 4D-Var minimizations at half the operational model resolution) with respect to the CNMCA analyses. The RMSE scores thus incorporate an intrinsic error of representativeness whose magnitude has not yet been determined.

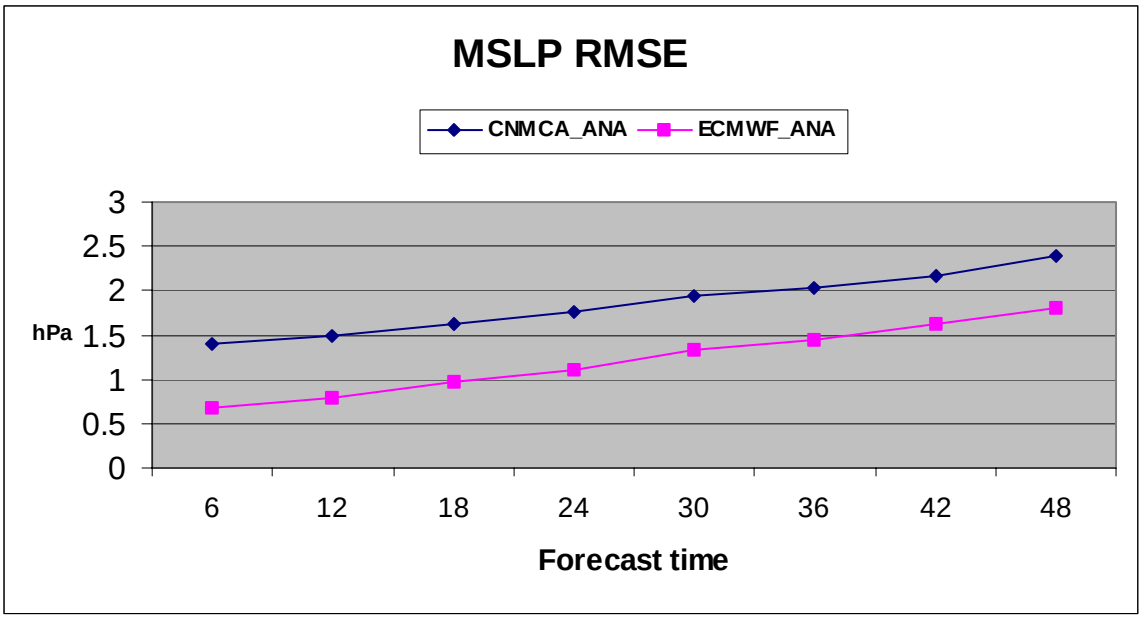

Figure 4.1 RMSE of Mean Sea level pressure forecast fields vs. ECMWF analysis.

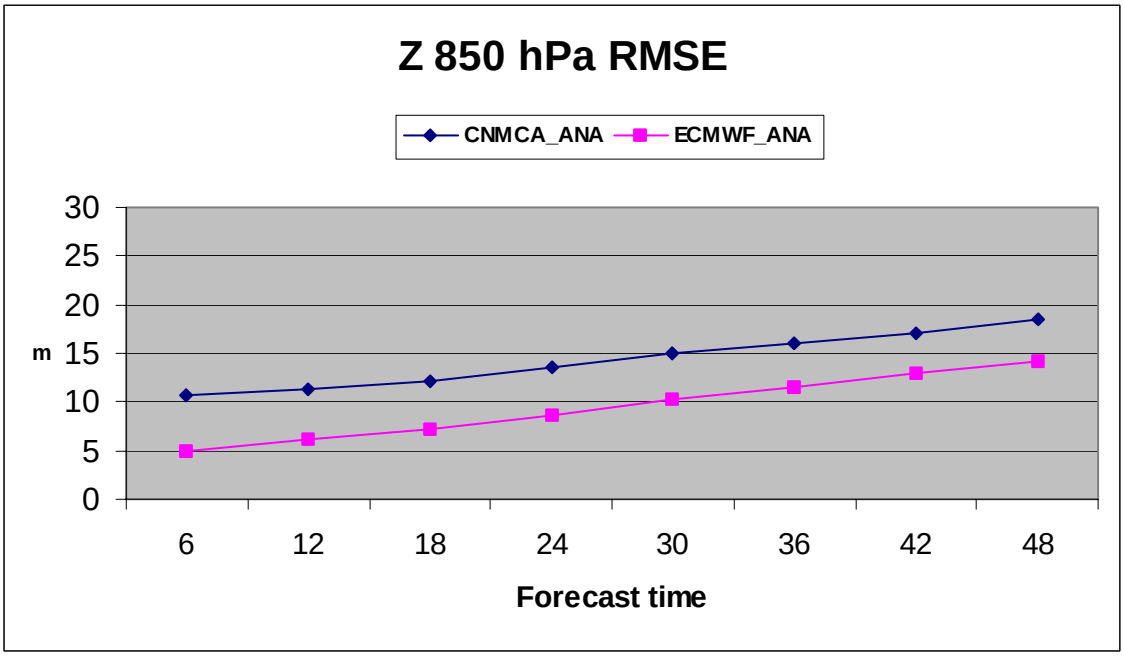

Figure 4.2 RMSE of 850 hPa geopotential height forecast fields vs. ECMWF analysis.

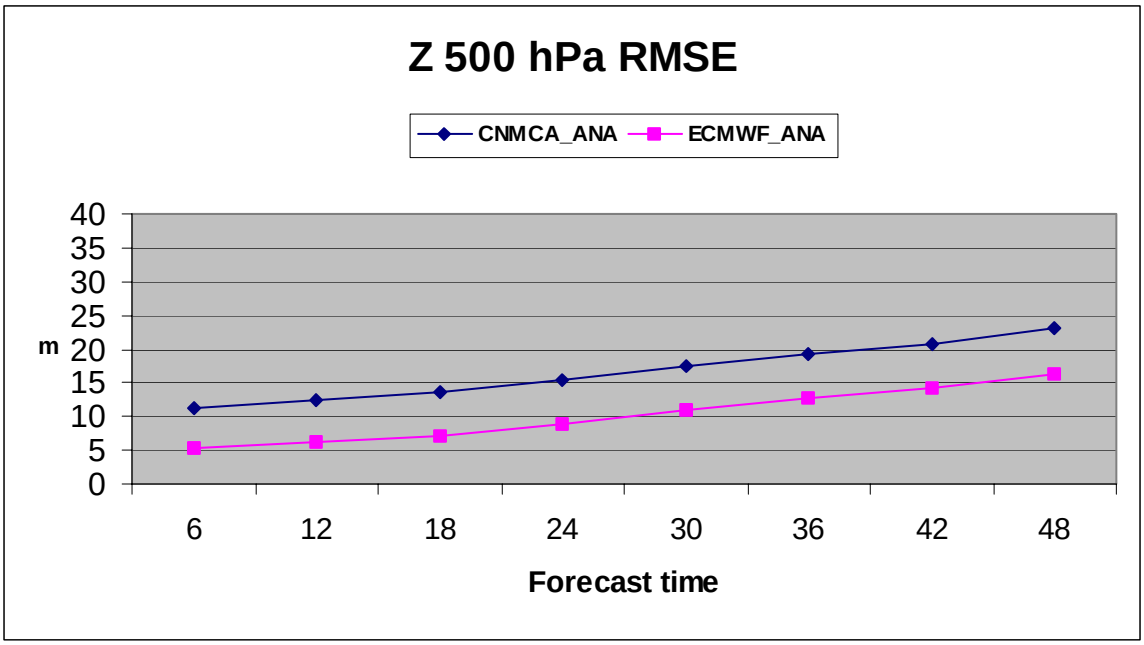

Figure 4.3 RMSE of 500 hPa geopotential height forecast fields vs. ECMWF analysis.

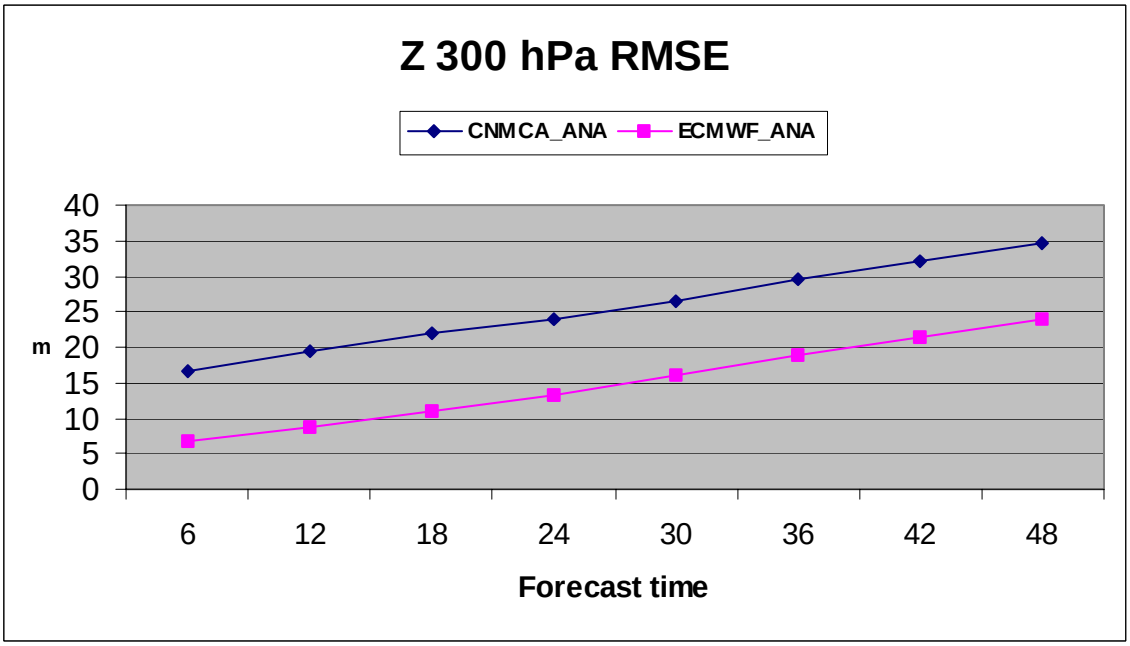

Figure 4.4 RMSE of 300 hPa geopotential height forecast fields vs. ECMWF analysis.

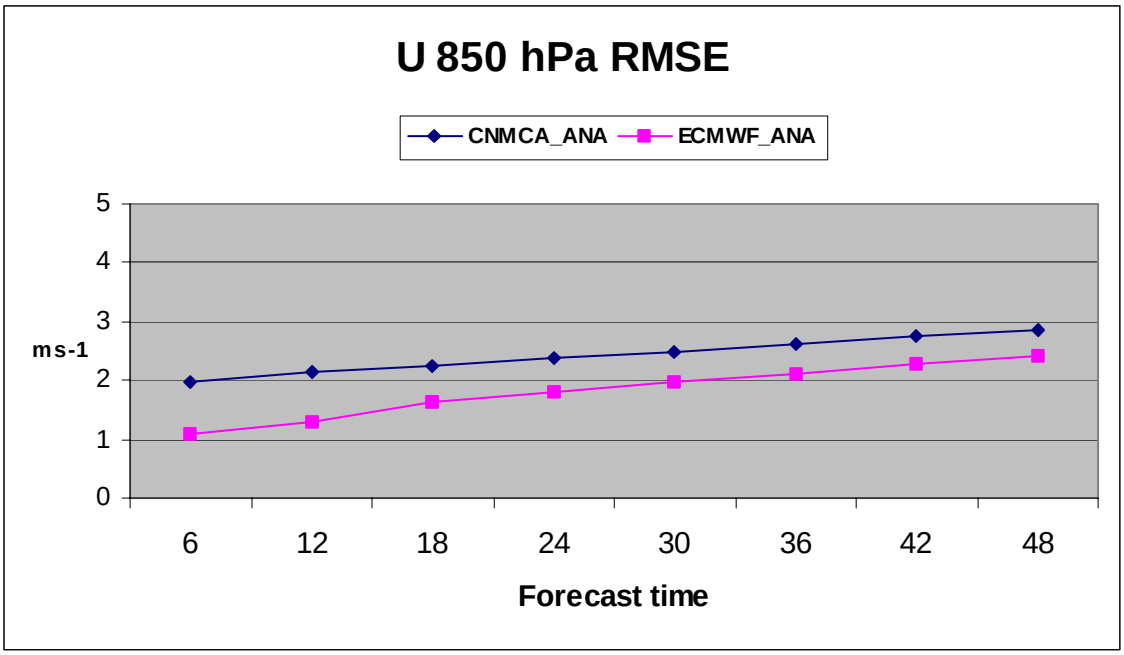

Figure 4.5 RMSE of 850 hPa wind speed forecast fields vs. ECMWF analysis.

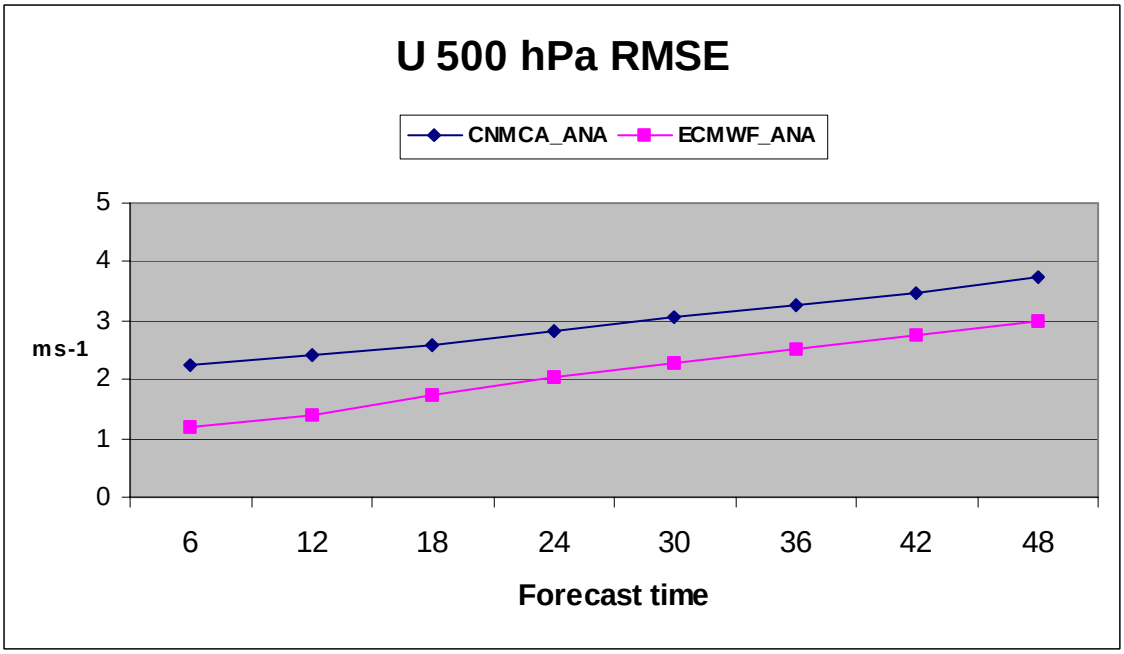

Figure 4.6 RMSE of 500 hPa wind speed forecast fields vs. ECMWF analysis.

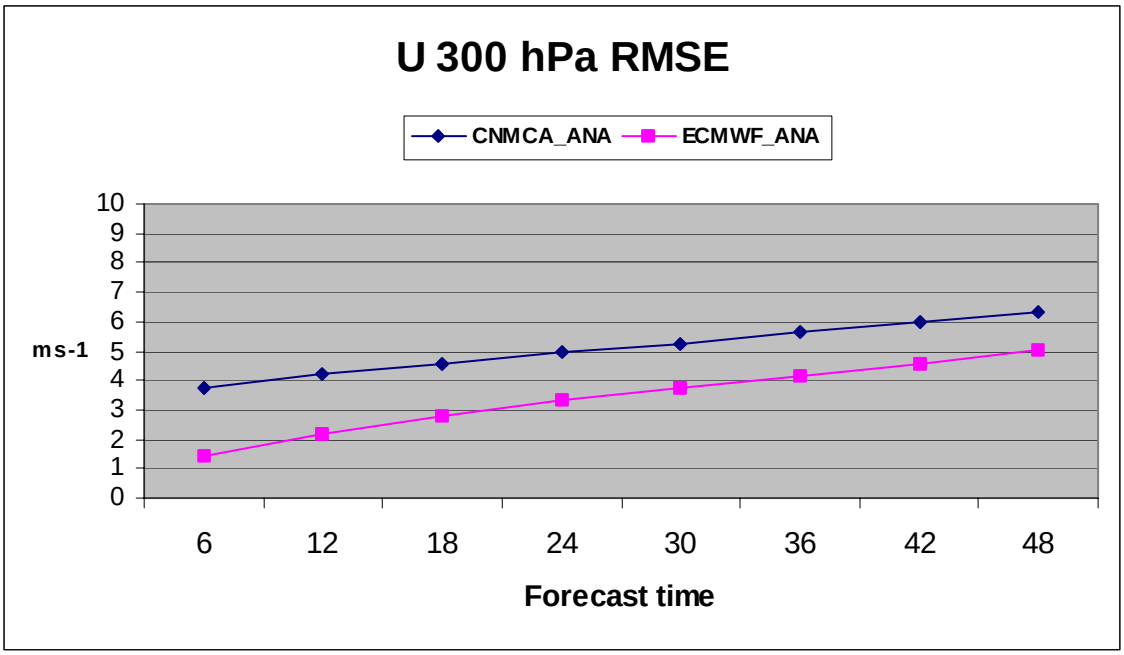

Figure 4.7 RMSE of 300 hPa wind speed forecast fields vs. ECMWF analysis.

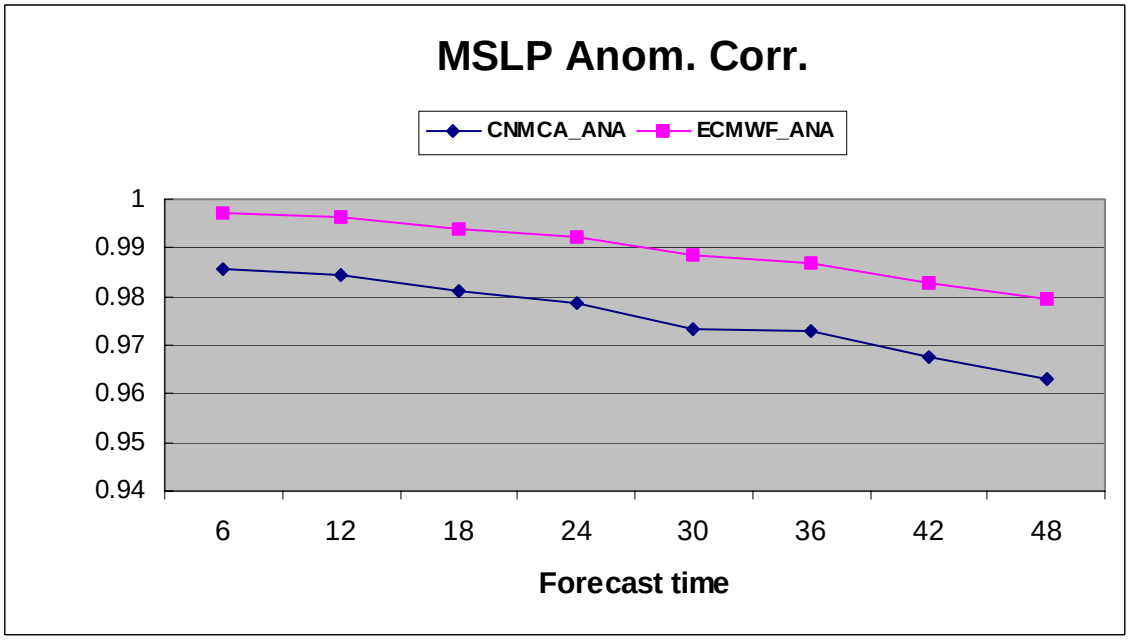

Figure 4.8 Anomaly Correlation of MSLP forecast fields vs. ECMWF analysis.

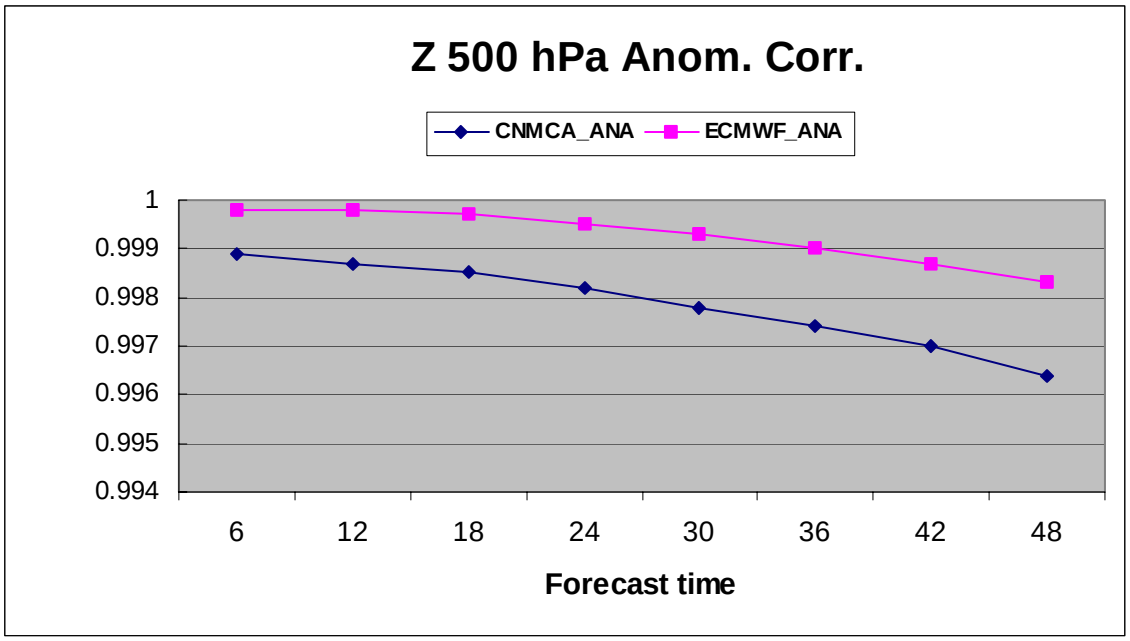

Figure 4.9 Anomaly Correlation of 500 hPa Geopotential height forecast fields vs. ECMWF analysis.

V. CONCLUSIONS AND SUGGESTIONS FOR FURTHER DEVELOPMENT

In this work the main ideas behind the new data assimilation cycle being implemented at Italian Air Force Weather Service have been presented. The system is in a relatively early stage of development and many important components are still missing in order to make it truly effective and operationally viable. However it is felt that the basic building blocks of the system have been set up:

1. The pre-processing and quality control of the observations;
2. The new 3-D covariance model in spherical coordinates;
3. The descent algorithm for the minimization of the cost function;
4. The statistical evaluation of the background error covariances.

At this point, the development effort will concentrate mainly on the following areas, which are deemed to be the most urgent requirements for an operational use of the new system:

1. The implementation of an effective buddy check algorithm for the screening of marginal observations. Given the global nature of the analysis scheme, a local method such as that first proposed by Lorenc (1981) would not be appropriate. Approximations to modern methods following Daley and Barker (2000, Section 9.3) are being investigated.

2. The implementation of an effective and easy to compute preconditioner for the minimization of the cost function is also being pursued. Currently the time required for the minimization algorithm is within acceptable limits, but it is expected that the inclusion of a larger number of satellite observations will change this.

3. The extension of the algorithm to observations that are not linearly related to the state variables. This work is well under way for scatterometer winds and will be started for column precipitable water observations and satellite radiances. The direct assimilation of satellite radiances is a long term goal whose main value is not expected to be so much in improving the current analyses over the present analysis domain as in the experience to be gained in view of the future availability of much improved observations from satellite hyperspectral sounders.

The results shown in the previous section indicate that the skill of the forecast fields derived from the new assimilation cycle still lags the skill of ECMWF derived forecasts by 18-24 h when verified with respect to ECMWF analyses. As mentioned in an earlier section, part of this is due to representativeness error. It is felt that this error can be reduced within 6-12h when more observations will have been added and some parameter optimizations will have been performed.

LIST OF REFERENCES

Benjamin, S. G. (1989): "An Isentropic Mesoα-scale Analysis System and its Sensitivity to Aircraft and Surface Observations", *Mon. Weath. Rev.*, Vol. **117**:1586-1603.

Bergman, K. H. (1979): "Multivariate Analysis of Temperatures and Winds Using Optimum Interpolation", *Mon. Weath. Rev.*, Vol. **107**:1423-1444.

Bergthorsson, P., and Doos, B. (1955): "Numerical Weather Map Analysis", Tellus, Vol. **7**:329-340.

Bjerknes, V. (1911): "Dynamical Meteorology and Hydrography. Part II. Kinematics", New York: Carnegie Institute, Gibson Bros.).

Bratseth, A. (1986): "Statistical Interpolation by Means of Successive Corrections", *Tellus,* Vol. **38A**: 439-47.

Brown, R.A. and G. Levy (1986): "Ocean Surface Pressure Fields from Satellite-Sensed Winds". *Mon. Wea. Rev*., Vol. **114**: 2197-2206.

Cohn, S. E. (1997): "An Introduction to Estimation Theory", *J. Meteor. Soc. Japan.* Special Volume "Data Assimilation in Meteorology and Oceanography: Theory and Practice" Vol. **75**, No **1B**, 257-88.

Cohn, S. E., Da Silva, A., Guo, J., Sienkiewicz, M. and D. Lamich (1998): "Assessing the Effects of Data Selection with the DAO Physical-Space Statistical Analysis System ". *Mon. Wea. Rev*., Vol. **126**: 2913-26.

Courtier, P. (1997): "Dual Formulation of Four-Dimensional Variational Assimilation", *Quart. J. Roy. Meteor. Soc.* Vol. **124**, 2449-61.

Courtier, P. (1997): "Variational Methods", *J. Meteor. Soc. Japan.* Special Volume "Data Assimilation in Meteorology and Oceanography: Theory and Practice" Vol. **75**, No **1B**, 319-29.

Courtier, P. and Coauthors (1998): "The ECMWF Implementation of Three-Dimensional Variational Assimilation (3D-VAR): Part I: Formulation", *Quart. J. Roy. Meteor. Soc.* Vol. **124**, 1783-1808.

Daley, R. (1983): "Spectral Characteristics of the ECMWF Objective Analysis System", ECMWF Technical Report N. 40 (Reading, England: ECMWF).

Daley, R. (1985): "The Analysis of Synoptic Scale Divergence by a Statistical Interpolation Procedure", *Mon. Weath. Rev.,* Vol. **113**:1066-1079.

Daley, R. (1991): "Atmospheric Data Analysis". Cambridge University Press, 458pp.

Daley, R. (1997): "Atmospheric Data Assimilation", *J. Meteor. Soc. Japan.* Special Volume "Data Assimilation in Meteorology and Oceanography: Theory and Practice" Vol. **75**, No **1B**, 319-29.

Daley, R. and Barker, E. (2000): "NAVDAS Source Book 2000". NRL Publication, NRL/PU/7530-00-418. Naval Research Laboratory, Marine Meteorology Division, Monterey, CA (USA).

Dévényi, D. and T.W. Schlatter (1994): "Statistical Properties of Three-hour Prediction "Errors" Derived from the Mesoscale Analysis and Prediction System ". *Mon. Wea. Rev*., Vol. **122**: 1263-1280.

ECMWF: (2001) "ECMWF Global Data Monitoring Report, November 2001", ECMWF Shinfield Park, RG2 9AX, Reading, Berkshire, England.

Evensen G. (199$) "Sequential Data Assimilation with a Nonlinear Quasi-geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics". *J. Geophys. Res*., 99(C5), 10143-10162.

Fisher M. (1998) “Development of a Simplified Kalman Filter”, ECMWF Technical memorandum 260.

Fisher M. (2001) “Assimilation Techniques: 3dVar”, ECMWF Training Course Notes: Data Assimilation and the use of Satellite data.

Gandin, L.S., (1963): "Objective Analysis of Meteorological Fields". Israeli Program for Scientific Translations, 1965, 242pp.

Golub, G. and H. van Loan, (1996): "Matrix Computations, third edition”. The John Hopkins University Press, 694 pp.

Haltiner, G. and Williams, R.T. (1980): "Numerical Prediction and Dynamic Meteorology". J. Wiley & Sons, 477pp.

Hollingsworth, A., and P. Lonnberg, (1986): “The Statistical Structure of Short-range Forecast Errors as Determined from Radiosonde Data. Part I: The Wind Field.” *Tellus*, **38A**, 111-136.

Kolmogorov, A., (1941): "Interpolated and Extrapolated Stationary Random Sequences". Izvestia an SSSR, seriya mathematicheskaya, Vol. **5(2)**: 85-95.

Ledvina, D. V. and Pfaendtner, J. (1995): "Inclusion of SSM/I total Precipitable Water Estimates into the GEOS-1 Data Assimilation System". *Mon. Wea. Rev*., Vol. **123**: 3003-3015.

Lorenc, A.C., (1981): "A Global Three-dimensional Multivariate Statistical Interpolation Scheme". *Mon. Wea. Rev*., Vol. **109**: 701-21.

Lorenc, A.C., (1986): "Analysis Methods for Numerical Weather Prediction". *Q.J.R. Meteorol. Soc*.,Vol. **112**,1177-94.

Louis, J. F., (**1979):** "A Parametric Model of Vertical Eddy Fluxes in the Atmosphere." *Bound.-Layer Meteor.,* Vol. **17,** 187-202

Majewski, D. (2001): "HRM User's Guide", available from the author (detlev.majewski@dwd.de)

Menard, R. (1994): "Kalman Filtering of Burgers' Equation and its Application to Atmospheric Data Assimilation", *PhD Thesis, McGill University, Stormy Weather Group Scientific Report, MW-100.*

Mellor, G.L. and T. Yamada (1974), "A Hierarchy of Turbulence Closure Models for Planetary Boundary Layers", *J. Atmos. Sci*., Vol. **31**, 1791-1806.

Miller, P.A. and S.G. Benjamin, (1992), "A System for the Hourly Assimilation of Surface Observations in Mountainous and Flat Terrain". *Mon. Wea. Rev*., Vol. **120**: 2342-2359.

Norris, B. (1990): "Pre-processing - General Data checking and validation". *ECMWF* Meteorological Bulletin, M1.4/3.

Otte, T.L., Seaman, N.L. and Stauffer, D.R. (2001): "A Heuristic Study on the Importance of Anisotropic Error Distributions in Data Assimilation". *Mon. Wea. Rev*., Vol. **129**: 766-83.

Parrish, D.F. and Derber, J.C. (1992): "The National Meteorological Center's Spectral Statistical Interpolation System". *Mon. Wea. Rev*., Vol. **120**: 1747-1763.

Rabier F. and P. Courtier, (1992), "Four Dimensional Assimilation in the Presence of Baroclinic Instability". *Q. J. R. Meteorol .Soc*., Vol. **118**, 649-672.

Rabier F. and Coauthors, (2000), "The ECMWF Operational Implementation of Four-Dimensional Variational Assimilation. I: Experimental Results with Simplified Physics". *Q. J. R. Meteorol. Soc*., Vol. **126**, 1143-1170.

Riishojgaard L.P., (1998), "A Direct Way of Specifying Flow Dependent Background Error for Meteorological Analysis Systems". *Tellus*, Vol. **50A**, 42-57.

Riishojgaard L.P., (2000), "A Method for Estimating the Analysis Error Variance in a Physical Space Data Assimilation System". *Q. J. R. Meteorol. Soc*., Vol. **126**, 1367-1386.

Ritter, B., and J.-F. Geleyn, (1992): "A Comprehensive Radiation Scheme for Numerical Weather Prediction Models with Potential Applications in Climate Simulations". *Mon. Wea. Rev.*, Vol. **120**, 303-325.

Rutherford, I. D. (1972): "Data Assimilation by Statistical Interpolation of Forecast Errors". *J.Atmos. Sci.*, Vol. **29**, 809.

Temperton, C. (1988), "Implicit Normal Mode Initialization". *Mon. Wea. Rev*, Vol. **116**, 1013-1031.

Thiebaux, H.J., Mitchell, H.L. and D.W. Shantz (1986), "Horizontal Structure of Hemispheric Forecast Error Correlations for Geopotential and Temperature". *Mon. Wea. Rev*, Vol. **114**, 1048-1066.

Thiebaux, H.J., Morone, L.L. and R.L. Wobus (1990), "Global Forecast Error Correlation. Part 1: Isobaric Wind and Geopotential". *Mon. Wea. Rev*, Vol. **118**, 2117-2137.

Tiedtke, M., (1989), "A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models." *Mon. Wea. Rev*., Vol. **117**, 1779-1800.

Tillman, G., (1999), "Correlations Functions for Atmospheric Data Analysis" *Q. J. R. Meteorol. Soc*., Vol. **125**, 2449-2464.

Vocino, A., (2002), "The diagnostic Component of the CNMCA Numerical Prediction System: Characteristics, Performance and Future Developments of the New Data Assimilation System" Thesis, 32nd Meteorology and Atmospheric Physics Course, CNMCA, Pratica di Mare, 90pp.(In Italian)

Wiener, N., (1949), "Extrapolation, Interpolation and Smoothing of Stationary Time Series" New York, John Wiley.

Wilkes, D.S. (1995). "Statistical Methods in the Atmospheric Sciences: An Introduction." Academic Press, 467 pp.

Xu, L. and R. Daley (2000), "Towards a True 4-dimensional Data Assimilation Algorithm: Application of a Cycling Representer Algorithm to a Simple Transport Problem" *Tellus,* 52A, 109-128.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Prof. Carlyle H. Wash
   Naval Postgraduate School
   Monterey, California

4. Prof. Roger T. Williams
   Naval Postgraduate School
   Monterey, California

5. Prof. Robert L. Haney
   Naval Postgraduate School
   Monterey, California

6. Dr. Nancy Baker
   Marine Meteorology Division.
   Naval Research Laboratory.
   Monterey, California

7. Gen. Roberto Sorani
   Ufficio Generale per la Meteorologia
   Aeroporto Centocelle, Rome, Italy

8. Col. Maurizio Bassani
   CNMCA
   Aeroporto De Bernardi, Pratica di Mare, Pomezia (RO), Italy

9. Col. Massimo Capaldo
   CNMCA
   Aeroporto De Bernardi, Pratica di Mare, Pomezia (RO), Italy

10. Ten.Col. Massimo Ferri
    CNMCA
    Aeroporto De Bernardi, Pratica di Mare, Pomezia (RO), Italy

11. Ms Annamaria Munno
    Office of Defense Cooperation, US Embassy in Italy
    Rome, Italy

12. Cap. Massimo Bonavita
CNMCA
Aeroporto De Bernardi, Pratica di Mare, Pomezia (RO), Italy